# Microlocal Analysis of the Geometric Separation Problem

David L. Donoho and Gitta Kutyniok

April 20, 2010

**Abstract**

Image data are often composed of two or more geometrically distinct constituents; in galaxy catalogs, for instance, one sees a mixture of pointlike structures (galaxy superclusters) and curvelike structures (filaments). It would be ideal to process a single image and extract two geometrically 'pure' images, each one containing features from only one of the two geometric constituents. This seems to be a seriously underdetermined problem, but recent empirical work achieved highly persuasive separations.

We present a theoretical analysis showing that accurate geometric separation of point and curve singularities can be achieved by minimizing the $\ell_1$ norm of the representing coefficients in two geometrically complementary frames: wavelets and curvelets. Driving our analysis is a specific property of the ideal (but unachievable) representation where each content type is expanded in the frame best adapted to it. This ideal representation has the property that important coefficients are *clustered geometrically* in phase space, and that at fine scales, there is very little coherence between a cluster of elements in one frame expansion and individual elements in the complementary frame. We formally introduce notions of *cluster coherence* and *clustered sparsity* and use this machinery to show that the underdetermined systems of linear equations can be stably solved by $\ell_1$minimization; microlocal phase space helps organize the calculations that cluster coherence requires.

**Key Words.** $\ell_1$ minimization. Sparse Representation. Mutual Coherence. Cluster Coherence. Tight Frames. Curvelets, Shearlets, Radial Wavelets.

**Acknowledgements.** The authors would like to thank Inam ur Rahman, Apple Computer, for graphics help, and Emmanuel Candès, Michael Elad, and Jean-Luc Starck, for numerous discussions on related topics. The second author would like to thank the Statistics Department at Stanford and the Mathematics Department at Yale for hospitality and support during her visits. Thanks also to the Isaac Newton Institute of Mathematical Sciences in Cambridge, UK for an inspiring research environment which led to the completion of a significant part of this work. This work was partially supported by NSF DMS 05-05303 and DMS 01-40698 (FRG), and by Deutsche Forschungsgemeinschaft (DFG) Heisenberg fellowship KU 1446/8-1.

# 1 Introduction

Cosmological data analysts face tasks of *geometric separation* [39, 40]. Gravitation, acting over time, drives an initially quasi-uniform distribution of matter in 3D to concentrate near lower-dimensional structures: points, filaments, and sheets. It would be desirable to process single 'maps' of matter density and somehow extract three 'pure' maps containing just the points, just the filaments, and just the sheets around which matter is concentrating.

In seemingly unrelated fields, such as medical imaging and materials science, related questions arise frequently and naturally. For example, a technologist with a single confused image of an aggregate might wish to create two images, one containing just the fibrous and the other just the granular structures, respectively.

Such 'desires', when voiced by a working scientist or engineer, really amount to a request for existing information technology to be put to work here and now on data available today. No doubt there is a wide spectrum of image processing 'hacks' and 'improvisations' that might be useful, on a case-by-case basis. The mathematician's interest would only be piqued when an intellectually coherent approach shows promise of success, especially if the reasons for success are subtle and instructive.

Recently, astronomer Jean-Luc Starck and collaborators have been empirically successful in numerical experiments with component separation; their approach used tools from modern harmonic analysis in a provocative way. They used two or more overcomplete frames, each one specially adapted to particular geometric structures, and were able to obtain separation despite the fact that the underlying system of equations is highly underdetermined. Here we analyze such approaches in a mathematical framework where we can show that success stems from an interplay between geometric properties of the objects to be separated, and the harmonic analysis for singularities of various geometric types. We eventually point to a much wider range of seemingly very different 'imaging' problems where our analysis techniques can provide insight.

## 1.1 Singularities and Sparsity

As a mathematical idealization of 'image', consider a Schwartz distribution $f$ with domain $\mathbf{R}^2$. The distribution $f$ will be given singularities with specified geometry: points and curves.

We plan to represent such an 'image' using tools of harmonic analysis; in particular, bases and frames. While many such representations are conceivable, we are interested here just in those bases or frames which can sparsely represent $f$ – i.e., can represent $f$ using relatively few large coefficients.

The type of basis which best sparsifies $f$ depends on the geometry of its singularities. If the singularities occur at a finite number of (variable) points, then *wavelets* give what is, roughly speaking, an optimally sparse representation – one with the fewest significantly nonzero coefficients. If the singularities occur at a finite number of smooth curves, then one of the recently studied directional multiscale representations (*curvelets* or *shearlets*) will do the best job of sparsification. (For careful quantitative discussions of sparsification see, e.g., [6] etc.).

In fact, real-world signals are, generally speaking, a mixture of content types and, cor-

respondingly, a model where singularities are of only one geometric type is overly narrow. If $f$ is actually a nontrivial superposition $\mathcal{P} + \mathcal{C}$ where $\mathcal{P}$ has only point singularities and $\mathcal{C}$ has only curvilinear singularities, then two things happen:

- Neither wavelets *alone* nor curvelets *alone* will be very good for representing $\mathcal{P} + \mathcal{C}$. The sparsity either achieves alone is much less satisfactory than the *ideal* sparsity level – that which could be achieved by using wavelets for representing $\mathcal{P}$ and by curvelets in representing $\mathcal{C}$. (This ideal representation is purely notional; it assumes one can first perfectly separate the two objects and then separately analyze the separated layers.)

- In fact, no single basis or traditional linear representation is very good at sparsifying $\mathcal{P} + \mathcal{C}$ compared to the ideal representation.

This immediately suggests the need to use *both* systems to represent $f$ sparsely; however, since each system is itself complete (or even overcomplete) there is no obvious traditional way to do this.

In this paper, we consider the problem of developing sparse representations by combining both wavelets and curvelets and using a nonlinear representation based on $\ell_1$ minimization. The problem we solve is a continuum variant of a problem in image and signal processing with considerable practical interest, and extensive work for almost two decades. For references, see Subsection 1.6 below. That work, while suggestive and inspiring, concerns discretely indexed signal/image processing, obscuring the *continuum* elements of geometry and microlocal analysis which are essential to this paper.

### 1.2 A Geometric Separation Problem

Consider the following simple but clear model problem of geometric separation. Consider a 'pointlike' object $\mathcal{P}$ made of point singularities:

$$\mathcal{P} = \sum_{i=1}^{P} |x - x_i|^{-3/2}. \tag{1.1}$$

This object is smooth away from the $P$ given points $(x_i : 1 \le i \le P)$. Consider as well a 'curvelike' object $\mathcal{C}$, a singularity along a closed curve $\tau : [0,1] \mapsto \mathbf{R}^2$:

$$\mathcal{C} = \int \delta_{\tau(t)}(\cdot) dt, \tag{1.2}$$

where $\delta_x$ is the usual Dirac delta function located at $x$. The singularities underlying these two distributions are geometrically quite different, but the exponent $3/2$ is chosen so the energy distribution across scales is similar; if $\mathcal{A}_r$ denotes the annular region $r < |\xi| < 2r$,

$$\int_{\mathcal{A}_r} |\hat{\mathcal{P}}|^2(\xi) \asymp r, \qquad \int_{\mathcal{A}_r} |\hat{\mathcal{C}}|^2(\xi) \asymp r, \qquad r \to \infty. \tag{1.3}$$

This choice makes the components comparable as we go to finer scales; the ratio of energies is more or less independent of scale. Separation is challenging at *every* scale.

Now assume that we observe the 'Signal'

$$f = \mathcal{P} + \mathcal{C}, \tag{1.4}$$

however, the component distributions $\mathcal{P}$ and $\mathcal{C}$ are unknown to us.

**Definition 1.1** *The* Geometric Separation Problem *requires to recover $\mathcal{P}$ and $\mathcal{C}$ from knowledge only of $f$; here $\mathcal{P}$ and $\mathcal{C}$ are unknown to us, but obey (1.1), (1.2) and certain regularity conditions on the curve $\tau$.*

As there are two unknowns ($\mathcal{P}$ and $\mathcal{C}$) and only one observation ($f$), the problem seems improperly posed. We develop a principled, rational approach which provably solves the problem according to clearly stated standards.

### 1.3 Two Geometric Frames

We now focus on two overcomplete systems for representing the object $f$:

- *Radial Wavelets* – a tight frame with perfectly isotropic generating elements.
- *Curvelets* – a highly directional tight frame with increasingly anisotropic elements at fine scales.

We pick these because, as is well known, point singularities are coherent in the wavelet frame and curvilinear singularities are coherent in the curvelet frame. In Section 8.1 we discuss other system pairs. For readers not familiar with frame theory, we refer to [11], where terms like 'tight frame' – a Parseval-like property – are carefully discussed.

The point- and curvelike objects we defined in the previous subsection are real-valued distributions. Hence, for deriving sparse expansions of those, we will consider radial wavelets and curvelets consisting of real-valued functions. So only angles associated with radians $\theta \in [0, \pi)$ will be considered, which later on we will, as is customary, identify with $\mathbf{P}^1$, the real projective line.

We now construct the two selected tight frames as follows. Let $W(r)$ be an 'appropriate' window function, where in the following we assume that $W$ belongs to $C^\infty(\mathbf{R})$ and is compactly supported on $[-2, -1/2] \cup [1/2, 2]$ while being the Fourier transform of a wavelet. For instance, suitably scaled Lemariè-Meyer wavelets possess these properties. We define *continuous radial wavelets* at scale $a > 0$ and spatial position $b \in \mathbf{R}^2$ by their Fourier transforms

$$\hat{\psi}_{a,b}(\xi) = a \cdot W(a|\xi|) \cdot \exp\{ib'\xi\}.$$

The *wavelet tight frame* is then defined as a sampling of $b$ on a series of regular lattices $\{a_j \mathbf{Z}^2\}$, $j \geq j_0$, where $a_j = 2^{-j}$, i.e., the radial wavelets at scale $j$ and spatial position $k = (k_1, k_2)'$ are given by the Fourier transform

$$\hat{\psi}_\lambda(\xi) = 2^{-j} \cdot W(|\xi|/2^j) \cdot \exp\{ik'\xi/2^j\},$$

where we let $\lambda = (j, k)$ index position and scale.

For the *same* window function $W$ and a 'bump function' $V$, we define *continuous curvelets* at scale $a > 0$, orientation $\theta \in [0, \pi)$, and spatial position $b \in \mathbf{R}^2$ by their Fourier transforms

$$\hat{\gamma}_{a,b,\theta}(\xi) = a^{\frac{3}{4}} \cdot W(a|\xi|)V(a^{-1/2}(\omega - \theta)) \cdot \exp\{ib'\xi\}.$$

See [7] for more details. The *curvelet tight frame* is then (essentially) defined as a sampling of $b$ on a series of regular lattices

$$\{R_{\theta_{j,\ell}} D_{a_j} \mathbf{Z}^2\}, \qquad j \geq j_0, \quad \ell = 0, \ldots, 2^{\lfloor j/2 \rfloor} - 1, \tag{1.5}$$

where $R_\theta$ is planar rotation by $\theta$ radians, $a_j = 2^{-j}$, $\theta_{j,\ell} = \pi\ell/2^{j/2}$, $\ell = 0, \ldots, 2^{j/2} - 1$, and $D_a$ is anisotropic dilation by $diag(a, \sqrt{a})$, i.e., the curvelets at scale $j$, orientation $\ell$, and spatial position $k = (k_1, k_2)$ are given by the Fourier transform

$$\hat{\gamma}_\eta(\xi) = 2^{-j\frac{3}{4}} \cdot W(|\xi|/2^j)V((\omega - \theta_{j,\ell})2^{j/2}) \cdot \exp\{i(R_{\theta_{j,\ell}} D_{2^{-j}} k)'\xi\},$$

where let $\eta = (j, k, \ell)$ index scale, orientation, and scale. (For a precise statement, see [8, Section 4.3, pp. 210-211]).

Roughly speaking, the radial wavelets are 'radial bumps' with position $k/2^j$ and scale $2^{-j}$, while the curvelets live on anisotropic regions of width $2^{-j}$ and length $2^{-j/2}$. The wavelets are good at representing point singularities while the curvelets are good at representing curvilinear singularities.

Using the *same* window $W$, we can construct a family of filters $F_j$ with transfer functions

$$\hat{F}_j(\xi) = W(|\xi|/2^j), \qquad \xi \in \mathbf{R}^2.$$

These filters allow us to decompose a function $f$ into pieces $f_j$ with different scales, the piece $f_j$ at subband $j$ arises from filtering $f$ using $F_j$:

$$f_j = F_j \star f;$$

the Fourier transform $\hat{f}_j$ is supported in the annulus with inner radius $2^{j-1}$ and outer radius $2^{j+1}$. Because of our assumption on $W$, we can reconstruct the original function from these pieces using the formula

$$f = \sum_j F_j \star f_j, \qquad f \in L^2(\mathbf{R}^2).$$

The tight frames of curvelets and radial wavelets discussed above interact in a very local way with the filtering $F_j$.

**Lemma.** *Let $\mathcal{F}_j$ denote the range of the operator of convolution with $F_j$. Then curvelets at level $j'$ are orthogonal to $\mathcal{F}_j$ unless $|j' - j| \leq 1$. Similarly, radial wavelets at level $j'$ are orthogonal to $\mathcal{F}_j$ unless $|j' - j| \leq 1$.*

**Proof.** Indeed, $\mathcal{F}_j$ is the collection of all functions $f_j$ whose Fourier transform is representable as $\hat{f}_j(\xi) = W(|\xi|/2^j)\hat{f}(\xi)$ where $f \in L^2(\mathbf{R}^2)$. The support in frequency space of elements of $\mathcal{F}_j$ is thus an annulus $\mathcal{A}_j$ (say). The annuli have disjoint interiors if $|j - j'| > 1$. Hence $\mathcal{F}_j \perp \mathcal{F}_{j'}$ if $|j' - j| > 1$.

However, both the radial wavelet frame elements and the curvelet frame elements at level $j'$ belong to $\mathcal{F}_{j'}$. □

For future use, let $\Lambda_j$ denote the collection of indices $(j,k)$ of wavelets at level $j$, and

$$\Lambda_j^{\pm n} = \bigcup_{j'=j-n}^{j+n} \Lambda_{j'}.$$

Similarly, let $\Delta_j$ denote the indices $\eta = (j,k,\ell)$ of curvelets at level $j$, and let

$$\Delta_j^{\pm n} = \bigcup_{j'=j-n}^{j+n} \Delta_{j'}.$$

We conclude that elements of $\mathcal{F}_j$ can be represented using either only radial wavelets $\{\psi_\lambda : \lambda \in \Lambda_j^{\pm 1}\}$ or only curvelets $\{\gamma_\eta : \eta \in \Delta_j^{\pm 1}\}$.

## 1.4 Separation via $\ell_1$ Minimization

### 1.4.1 Sparse Multiple Frame Expansions

We now have two complete representations for $\mathcal{F}_j$, yielding two ways of representing the subband component $f_j$: in terms of its wavelet expansion:

$$f_j = \sum_{\lambda \in \Lambda_j^{\pm 1}} w_\lambda \psi_\lambda;$$

or in terms of its curvelet expansion:

$$f_j = \sum_{\eta \in \Delta_j^{\pm 1}} c_\eta \gamma_\eta.$$

Each frame exhibits a single geometric tendency – either highly nondirectional or highly directional – in representing $f_j$. However, $f_j$ may have both isotropic and directional features. We therefore seek a combined representation

$$f_j = \sum_{\lambda \in \Lambda_j^{\pm 1}} w_\lambda \psi_\lambda + \sum_{\eta \in \Delta_j^{\pm 1}} c_\eta \gamma_\eta.$$

Because the combined frame formed by concatenating the two frames is overcomplete, there are many possible ways this decomposition can be done. Some of them may be geometrically motivated, many are not.

Consider the following dual-frame **Component Separation** problem based on $\ell_1$ minimization:

$$\begin{aligned} (\text{CSEP}) \qquad (W_j, C_j) \;=\; & \operatorname{argmin} \|w\|_1 + \|c\|_1 \\ & \text{subject to } f_j = W_j + C_j \\ & \quad \text{and } w_\lambda = \langle W_j, \psi_\lambda \rangle, \quad \lambda \in \Lambda_j^{\pm 1} \\ & \quad \text{and } c_\eta = \langle C_j, \gamma_\eta \rangle, \quad \eta \in \Delta_j^{\pm 1}. \end{aligned}$$

In words, we take a given scale subband $f_j$ and decompose it into a wavelet component $W_j$ and a curvelet component $C_j$. The components are chosen by the principle of $\ell_1$ minimization on the frame coefficients: the $\ell_1$ norm of the wavelet coefficients of the wavelet component should be small, and the $\ell_1$ norm of the curvelet coefficients of the curvelet component should be small.

Here is our reason for the 'component separation' label: Armed with the optimization result at each scale subband, we define the purported *pointlike* component as the superposition of all the wavelet terms:

$$\tilde{P} = \sum_j F_j \star W_j;$$

and the purported *curvelike* component as the superposition of all the curvelet terms:

$$\tilde{C} = \sum_j F_j \star C_j.$$

We obtain the decomposition

$$f = \tilde{P} + \tilde{C}.$$

### 1.4.2 Main Result

At this stage, we have two decompositions: one by the truly geometric pair $(\mathcal{P}, \mathcal{C})$ of pointlike and curvelike objects and one by the *purported* geometric pair $(\tilde{P}, \tilde{C})$. The following result justifies our interest in the second pair. To state it, define the scale subbands of the truly geometric components by:

$$\mathcal{P}_j = F_j \star \mathcal{P}; \qquad \mathcal{C}_j = F_j \star \mathcal{C}.$$

**Theorem 1.1** Asymptotic Separation.

$$\frac{\|W_j - \mathcal{P}_j\|_2 + \|C_j - \mathcal{C}_j\|_2}{\|\mathcal{P}_j\|_2 + \|\mathcal{C}_j\|_2} \to 0, \qquad j \to \infty. \tag{1.6}$$

At fine scales, the truly pointlike component is almost all captured by the wavelet component and the truly curvelike component is almost all captured by the curvelet component. In short, the purported pointlike and curvelike components deserve the labelling they have been given.

## 1.5 Extensions

Theorem 1.1 is amenable to generalizations and extensions. Previewing Section 8.1, we mention a few examples.

- *More General Classes of Objects.* Theorem 1.1 can be generalized to other situations. First, we could consider singularities of different orders. This would allow $\mathcal{C}$ to model 'cartoon' images, where the curvilinear singularities are now the boundaries of the pieces for piecewise $C^2$ functions. Second, we can allow smooth perurbations, i.e., $f = (\mathcal{P} + \mathcal{C} + g) \cdot h$ where $g, h$ are smooth functions of rapid decay at $\infty$. In this situation, we let the denominator in (1.6) be simply $\|f_j\|_2$.

- *Other Frame Pairs.* Theorem 1.1 holds without change for many other pairs of frames and bases, such as, e.g., orthonormal separable Meyer wavelets and shearlets.
- *Noisy Data.* Theorem 1.1 is resilient to noise impact; an image composed of $\mathcal{P}$ and $\mathcal{C}$ with additive 'sufficiently small' noise exhibits the same asymptotic separation.
- *Rate of Convergence.* Theorem 1.1 can be accompanied by explicit decay estimates.
- *Other Algorithms and Other Notions of Separation.* In the companion paper [20] we study thresholding as an alternative approach to separation; it is less computationally demanding than the $\ell^1$ minimization studied here, but also somewhat less elegant. Building on the estimates proved in this paper, [20] shows that properly-tuned thresholding can also achieve asymptotic separation.

### 1.6 The Multiple-Basis Representation Problem

Theorem 1.1 should be placed in context of a great deal of ongoing work concerning sparsity and overcomplete representations. Already in the early 1990's, R.R. Coifman became interested in the problem of representing discrete-time signals using more than one basis. In a conversation, he told one of us about a problem which, in retrospect and using modern formulations, can be posed as follows:

- An observed signal $S \in \mathbf{R}^n$ is thought to be a superposition of subsignals $S_i$, $i = 1, 2$.
- Each subsignal $S_i$ is thought to be 'coherent' in an 'appropriate' basis $\Phi_i$, $i = 1, 2$.
- Each subsignal 'looks incoherent' in an 'inappropriate' basis. Here $\Phi_2$ is inappropriate for $S_1$, and $\Phi_1$ is inappropriate for $S_2$.

Coifman, Wickerhauser and co-workers at the time made a sort of heuristic exploration motivated intuitively by these slogans. As a published example of their work at the time, please see [12, Fig. 26(a-h)]. The different 'coherent parts' displayed in those figures were obtained by the following recipe:

1. Transform signal $S$ into basis $\Phi_1$.
2. Threshold the coefficients, yielding sparse coefficients $\tilde{\alpha}_1$.
3. Form residual $R = S - \Phi_1 \tilde{\alpha}_1$.
4. Transform $R$ into basis $\Phi_2$.
5. Threshold the coefficients, yielding sparse coefficients $\tilde{\alpha}_2$.
6. Write $\tilde{S}_i = \Phi_i \tilde{\alpha}_i$; then
$$S = \tilde{S}_1 + \tilde{S}_2 + residual. \tag{1.7}$$

At about the same time, Stéphane Mallat and Zhifeng Zhang became interested in the problem of representing signals using a highly overcomplete dictionary of time-frequency atoms; [34] (their dictionary had $\approx \log(N)$ different frames, where $N$ is the signal length). Their approach, called *Matching Pursuit*, built up an approximation one-term-at-a-time iteratively, at each stage finding the best single atom in any of the several bases which was not yet already forming part of the approximation and adding that term to the approximation.

Lurking in these early numerical experiments were two larger questions. If there truly is a simple representation of the signal using more than one system, can it ever be found? Can it be found by such a simple approach?

Formally: can one accurately recover 'coherent' pieces $S_1$ and $S_2$ given knowledge of $S = S_1 + S_2$ only? For example, can we expect that the outputs $\tilde{S}_1$, $\tilde{S}_2$ in (1.7) obey $\tilde{S}_1 \approx S_1$ and $\tilde{S}_2 \approx S_2$? Researchers at the time said in conversation, that, when put this starkly, the answer was simply 'no', since there are twice as many unknowns as knowns. Nevertheless, some of the empirical results at the time were suggestive and inspiring.

### 1.7 Minimum $\ell_1$ Decomposition and Perfect Separation

A few years later, one of us worked with Scott Shaobing Chen to develop a formal, optimization-based approach to the multiple-basis representation problem. Given bases $\Phi_i$, $i = 1, 2$, one solves the following problem

$$\text{(BP)} \qquad \min \|\alpha_1\|_1 + \|\alpha_2\|_1 \text{ subject to } S = \Phi_1\alpha_1 + \Phi_2\alpha_2.$$

Here $\|\cdot\|_1$ denotes the usual $\ell_1$ norm. Note that here there are $2n$ unknowns in $\alpha_1$ and $\alpha_2$ and only $n$ knowns in $S$, but that an optimization principle is being used to select a particular element from the $n$-dimensional space of all possible solutions. (Terminological note: the name 'Basis Pursuit' is meant to remind the reader that (BP) actually selects a basis for the solution out of the many conceivable bases which can be extracted from the union of the two overcomplete systems).

Based on earlier experience of our first-named author and his collaborators, see [19, 22, 23], it was known that the $\ell_1$ norm had a tendency to find sparse solutions when they exist. And indeed, Chen's thesis showed that in some simple special cases that this was so. Letting $\Phi_1$ be the standard basis of $\mathbf{R}^n$ (i.e., Kronecker sequences or 'spikes') and $\Phi_2$ be the Fourier basis, Chen considered signals $S$ which were superpositions of two spikes and two sinusoids. He showed that (BP) recovered exactly the indices and coefficients of the terms involved in the synthesis; and that this was true across a wide range of amplitude ratios between the sinusoid and spike components. In short, there was perfect separation of sinusoids from spikes, and the true underlying simplicity of the signal was revealed – even though there were more unknowns than equations.

In the years since that work, two streams of research emerged.

- *Theoretical work*, showing that, indeed, one could in certain settings obtain the sparsest possible representations to an underdetermined problem by $\ell_1$ optimization; see, e.g., [9, 13, 14, 15, 17, 45] for a selection of general work concerning $\ell_1$ minimization, and [2, 16, 18, 24, 27] for work somewhat relevant to component separation.

- *Empirical work*, showing that combined representations such as wavelets with curvelets or wavelets with sinusoids often gave very compelling separations of real signals and images, see, for instance, [1, 10, 25, 26, 35, 42, 40, 41, 43, 44, 30, 47].

We have already mentioned the empirical successes of Starck and collaborators. For an overview of much recent work on sparse decompositions, see [3]. Note that geometric separation is somewhat different from the task of separation of *texture from smooth structure*; in that problem, sparsity in frame expansions does not play an explicit role, nor do the geometric considerations which are so important here; very interesting early work in such non-geometric separation was published by Yves Meyer [36], and Vese and Osher [37].

### 1.8 Theorem 1.1 in Context, and Outline of Paper

We can now place our result in context, via several comparisons and contrasts, looking ahead to themes developed below.

- *Microlocal Viewpoint.* In Theorem 1.1 the objects of interest are collections of point and curve singularities. The viewpoint derives from microlocal analysis (see Section 3), which says that points and curves are very different objects in their joint space/orientation structure, so that even if they happen to overlap spatially, they are microlocally distinct. In contrast, other work on sparsity and $\ell_1$ minimization typically has a discrete flavor, making hypotheses about the number of nonzeros in an expansion and assuming the dictionary elements interact randomly.

- *Microlocal Asymptotics.* Asymptotics are important for Theorem 1.1; the sharp separation between curves and points in microlocal phase space exists only as a limit phenomenon, as the scale tends to zero. Asymptotic statements are important in other literature on sparsity-driven decompositions, but they are asymptotic in the number of random elements in the underlying matrix, and exploit law-of-large-numbers and concentration-of-measure effects. For Theorem 1.1, such principles play no role.

- *Clustered Sparsity.* In other work on sparsity-driven decompositions, sparsity of the coefficients plays a role primarily through the number of nonzeros. In this work sparsity plays a role also through the arrangement of nonzeros; Section 2.1 introduces a notion of clustered sparsity and Sections 4-7 develop estimates bounding the locations of significant nonzeros in the wavelet expansion of a point singularity or in the curvelet expansion of a curve singularity; the estimates will be organized using micolocal phase space ideas described in Section 3.

- *Cluster Coherence.* In other work on sparsity-driven decompositions, coherence or restricted isometry principles play a role; these don't depend on the arrangement of nonzeros in an expansion. Moreover, these are often applied to random-dictionary situations where the interaction between frame elements is random and quasi arbitrary. Section 2.1 develops the notion of *cluster coherence* which specifically depends on the arrangement of nonzeros. We apply this notion to a dictionary (wavelets + curvelets) where interactions are geometrically driven, and we develop estimates motivated by microlocal analysis which provide the needed geometrical information.

Our paper begins with Sections 2 and 3 introducing the driving ideas of clustered sparsity, cluster coherence, and microlocal separation, and describing a plan to prove Theorem 1.1 by establishing several needed estimates. Later Sections 4, 5, 6, and 7 then develop these estimates, by developing results about wavelet and curvelet expansions of point and line singularities. Section 8 then mentions a number of possible extensions.

# 2 Component Separation by $\ell_1$ Minimization

We now study the behavior of $\ell^1$ minimization in the two-frame case. Our analysis centers on the use of cluster coherence to control joint concentration.

## 2.1 $\ell_1$ Minimization for Separation of Two Tight Frames

Suppose we have two tight frames $\Phi_1$, $\Phi_2$ in a Hilbert space $\mathcal{H}$, and a signal vector $S \in \mathcal{H}$. We know *a priori* that there exists a decomposition

$$S = S_1^0 + S_2^0,$$

where $S_1^0$ is sparse in Frame 1, and $S_2^0$ is sparsely represented in Frame 2.

Consider the following optimization problem

$$\text{(Sep)} \qquad (S_1^\star, S_2^\star) = \text{argmin}_{S_1,S_2} \|\Phi_1^T S_1\|_1 + \|\Phi_2^T S_2\|_1 \text{ subject to } S = S_1 + S_2. \qquad (2.1)$$

The *optimization problem* (Sep) is visibly similar to, but subtly different from, (BP). Here the $\ell_1$ norm is being applied on the *analysis* coefficients of the two different 'components' rather than on the individual *synthesis* coefficients. The hope in (BP) is to get exactly the right nonzero coefficients in the sense of those providing the sparsest representation. However, this can become numerically unstable for certain tight frames. The hope in (Sep) is merely to separate components rather than the more ambitious goal of identifying the true nonzero coefficients within each component's representation. Starck and Elad have found this distinction to be important in their own empirical work on separation.

To analyze this we need the following notion.

**Definition 2.1** *Let $\Phi_1$ and $\Phi_2$ be two tight frames. Given two sets of coefficients $\mathcal{S}_1$ and $\mathcal{S}_2$, define the* joint concentration $\kappa = \kappa(\mathcal{S}_1, \mathcal{S}_2)$ *by*

$$\kappa(\mathcal{S}_1, \mathcal{S}_2) = \sup_f \frac{\|1_{\mathcal{S}_1} \Phi_1^T f\|_1 + \|1_{\mathcal{S}_2} \Phi_2^T f\|_1}{\|\Phi_1^T f\|_1 + \|\Phi_2^T f\|_1}.$$

In words, we consider the maximal fraction of total $\ell_1$ norm which can be concentrated to the combined index set $\mathcal{S}_1 \cup \mathcal{S}_2$. Concepts of this kind go back to [23]. Adequate control of joint concentration ensures that the principle (2.1) gives a successful approximate separation.

**Proposition 2.1** *Suppose that $S$ can be decomposed as $S = S_1^0 + S_2^0$ so that each component $S_i^0$ is relatively sparse in $\Phi_i$, $i = 1, 2$, i.e.,*

$$\|1_{\mathcal{S}_1^c} \Phi_1^T S_1^0\|_1 + \|1_{\mathcal{S}_2^c} \Phi_2^T S_2^0\|_1 \le \delta.$$

*Let $(S_1^\star, S_2^\star)$ solve (2.1). Then*

$$\|S_1^\star - S_1^0\|_2 + \|S_2^\star - S_2^0\|_2 \le \frac{2\delta}{1-2\kappa}.$$

**Definition 2.2** *Given tight frames $\Phi = (\phi_i)_i$ and $\Psi = (\psi_j)_j$ and an index subset $\mathcal{S}$ associated with expansions in frame $\Phi$, we define the* cluster coherence

$$\mu_c(\mathcal{S}, \Phi; \Psi) = \max_j \sum_{i \in \mathcal{S}} |\langle \phi_i, \psi_j \rangle|.$$

In many studies of $\ell_1$ optimization, one utilizes instead the mutual coherence

$$\mu(\Phi, \Psi) = \max_j \max_i |\langle \phi_i, \psi_j \rangle|, \tag{2.2}$$

whose importance was shown by [18]. This may be called the *singleton coherence.* In contrast, cluster coherence bounds coherence between a single member of frame $\Psi$ and a cluster of members of frame $\Phi$, clustered at $\mathcal{S}$.

A related notion called 'cumulative coherence' was introduced in [45]; that notion maximizes over subsets $\mathcal{S}$ of a given size, whereas here we fix a specific set $\mathcal{S}$ of coefficients. In applying our concept, the index subsets we will consider are not abstract, but will instead have a specific geometric interpretation, associated to proximity to certain curves in phase space. Maximizing over all subsets of a given size would give very loose bounds, and would not be suitable for our purposes. Several other coherence measures involving subsets appear in the literature, e.g., [4] and [46], but we do not see a strong relation to cluster coherence.

**Lemma 2.1** *We have*

$$\kappa(\mathcal{S}_1, \mathcal{S}_2) \le \max\{\mu_c(\mathcal{S}_1, \Phi_1; \Phi_2), \mu_c(\mathcal{S}_2, \Phi_2; \Phi_1)\}.$$

The proofs for Proposition 2.1 and Lemma 2.1 are presented in Section 9.1.

## 2.2 Intended Application

The concepts of this section will now be applied to (CSep), at scale $j$ only. With $f = \mathcal{P} + \mathcal{C}$ our distribution of interest, and $F_j$ our bandpass filter, and we set $f_j = F_j \star f$. Throughout this section, the object $S \equiv f_j$ and the tight frames are $\Phi_1$, the full radial wavelet frame, and $\Phi_2$, the full curvelet tight frame. We apply the optimization problem (Sep), getting subsignal components $S_1^\star$ and $S_2^\star$, which we then relabel as the wavelet component $W_j$ and curvelet component $C_j$; *one should check that with this sequence of substitutions, problem (*Sep*) of this section becomes (*CSep*) of the introduction.*

The key problem in the application of Proposition 2.1 to the aforementioned setting is the correct choice of the clusters of significant coefficients at each scale. If those clusters are chosen 'too small asymptotically', the relative sparsity will blow up, and if chosen 'too large asymptotically', we lose control of the cluster coherence. We define those clusters on the ideal decomposition, where wavelets are used to analyze the point singularity and curvelets are used to analyze the curve singularity. (Such clusters are theoretical, non-observable entities.)

For a series of wavelet-coefficient thresholds $\varepsilon_{j,1}$ to be specified, the cluster of significant wavelet coefficients can be provisionally defined as

$$\mathcal{S}_{1,j} = \{\lambda \in \Lambda_j^{\pm 1} : |\langle \psi_\lambda, \mathcal{P}_j \rangle| > \varepsilon_{j,1} \cdot \|(\langle \psi_{\lambda'}, \mathcal{P}_j \rangle)_{\lambda'}\|_{\ell_\infty(\Lambda_j^{\pm 1})}\}. \tag{2.3}$$

For a series of curvelet-coefficient thresholds $\varepsilon_{j,2}$ to be specified, the cluster of significant curvelet coefficients can be provisionally taken as

$$\mathcal{S}_{2,j} = \{\eta \in \Delta_j^{\pm 1} : |\langle \gamma_\eta, \mathcal{C}_j \rangle| > \varepsilon_{j,2} \cdot \|(\langle \gamma_{\eta'}, \mathcal{C}_j \rangle)_{\eta'}\|_{\ell_\infty(\Delta_j^{\pm 1})}\}. \tag{2.4}$$

Each threshold choice picks a specific point on the tradeoff between relative sparsity and cluster coherence.

Then $\delta_j$ will denote the degree of approximation by significant coefficients, the sum $\delta_j = \delta_{j,1} + \delta_{j,2}$ of the wavelet approximation error to the point singularity:

$$\delta_{j,1} = \sum_{\lambda \in \mathcal{S}_{1,j}^c} |\langle \psi_\lambda, \mathcal{P}_j \rangle|;$$

and the curvelet approximation error to the curvilinear singularity:

$$\delta_{j,2} = \sum_{\eta \in \mathcal{S}_{2,j}^c} |\langle \gamma_\eta, \mathcal{C}_j \rangle|.$$

Finally, $\kappa(\mathcal{S}_{1,j}, \mathcal{S}_{2,j})$, the degree of joint wavelet-curvelet concentration at the significant subsets, will be controlled by two cluster coherences: $\mu_c(\mathcal{S}_{1,j}, \Phi_1; \Phi_2)$ the maximal coherence of a curvelet to a cluster of significant wavelet coefficients; and $\mu_c(\mathcal{S}_{2,j}, \Phi_2; \Phi_1)$ the maximal coherence of a wavelet to a cluster of significant curvelet coefficients. We have

**Corollary 2.1** *Suppose that the sequence of transform-space clusters* $(\mathcal{S}_{1,j})$*, and* $(\mathcal{S}_{2,j})$ *has* **both** *of the following two properties: (i) asymptotically negligible cluster coherences:*

$$\mu_c(\mathcal{S}_{1,j}, \Phi_1; \Phi_2), \mu_c(\mathcal{S}_{2,j}, \Phi_2; \Phi_1) \to 0, \qquad j \to \infty,$$

*and (ii) asymptotically negligible cluster approximation errors:*

$$\delta_j = \delta_{1,j} + \delta_{2,j} = o(\|f_j\|_2), \qquad j \to \infty.$$

*Then we have asymptotically near-perfect separation:*

$$\frac{\|W_j - \mathcal{P}_j\|_2 + \|C_j - \mathcal{C}_j\|_2}{\|f_j\|_2} \to 0, \qquad j \to \infty.$$

The main result –Theorem 1.1 – follows from this lemma, but this will require sufficiently good estimates for cluster coherence for clusters defined as sufficiently good approximations to the objects of interest.

We remark that although the threshold in the *provisional* definition of the clusters in (2.3)-(2.4) provides a means to balance between relative sparsity and cluster coherence, there is no *a priori* guarantee that there exists a threshold for which conditions (i) and (ii) of Corollary 2.1 are true. The main achievement of this paper is to show that this is indeed possible with a revised definition. In fact we do not finally define the clusters by (2.3)-(2.4); note that the clusters must simply exhibit properties (i) and (ii) of the Corollary. We intend to make use of our freedom of definition in order to get the needed properties.

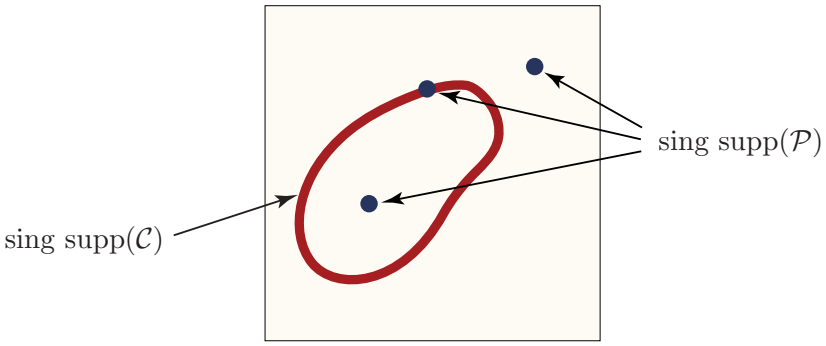


Figure 1: Singular supports of the point singularity $\mathcal{P}$ and the curvilinear singularity $\mathcal{C}$. The two supports overlap in one point.

# 3 Microlocal Analysis Viewpoint

The proof of the main result boils down to defining clusters and bounding cluster coherences and cluster approximation errors; our approach is inspired by microlocal analysis. In effect, we consider the case where the cluster is either a string of curvelet coefficients in the cone of influence of a curvilinear singularity, or a block of wavelet coefficients in the cone of influence of a point singularity, and we must bound interactions between clusters in one frame and elements in the other frame. Microlocal analysis provides a simple organizational framework that immediately suggests which interactions 'ought' to be small and large, based on the geometry of the overlaps between phase portraits in the underlying microlocal phase space.

## 3.1 Microlocal Analysis Concepts

The singular support of a distribution $f$, sing supp$(f)$, is the set of points where $f$ is not locally $C^\infty$. In the geometric separation setting, we have

$$\text{sing supp}(f) = \text{sing supp}(\mathcal{P} + \mathcal{C}) = \text{sing supp}(\mathcal{P}) \cup \text{sing supp}(\mathcal{C}) = \{x_i\} \cup image(\tau)$$

because we have constructed the distributions $\mathcal{P}$ and $\mathcal{C}$ so their singularities have this form.

Note that the points $x_i$ can intersect the image of the curve $\tau$ – we make no separation hypothesis asking the point singularities to 'stay away' from the curvilinear singularities. Figure 1 displays the singular support of $f$ and the contributions from $\mathcal{P}$ and $\mathcal{C}$.

To properly separate between pointlike and curvelike singularities we need to consider a *phase space* for microlocal analysis indexed by position-orientation pairs $(b, \theta)$; such pairs can describe the locations and orientations where $f$ has singular behavior. The orientational component $\theta$ will be regarded as an element in $\mathbf{P}^1$, the real projective space in $\mathbf{R}^2$. (Here we identify $\mathbf{P}^1$ with $[0, \pi)$ and freely write one or the other in what follows. It may at first seem more natural to think of directions $[0, 2\pi)$ rather than orientations $[0, \pi)$, note however that in this paper we consider *real-valued* distributions $\mathcal{P} + \mathcal{C}$ measured by real-valued curvelets $\gamma_\eta$ so directions are not resolvable, only orientations. We also frequently abuse notation as follows: we will write $|\theta - \theta'|$ when what is actually meant is geodesic distance between two points on $\mathbf{P}^1$.)

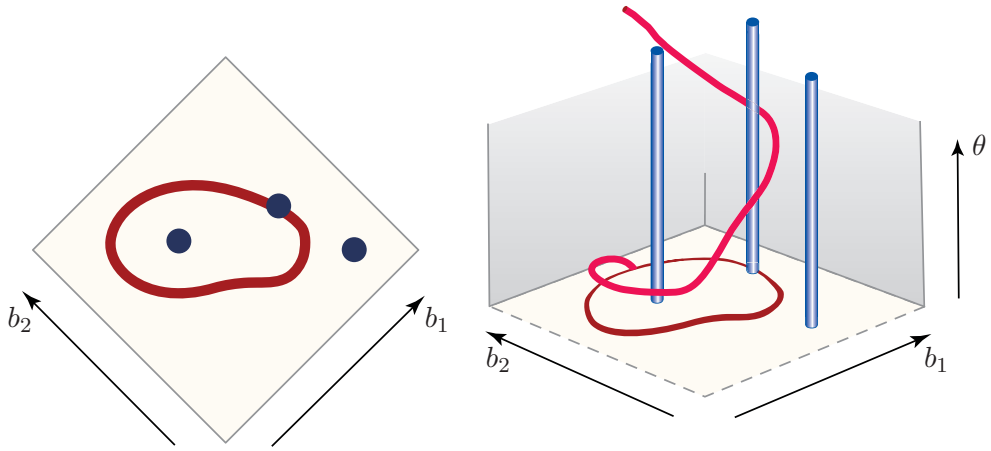


Figure 2: Left panel: singular supports of $\mathcal{P}$ and $\mathcal{C}$ (compare Figure 1). Right panel: wavefront sets of $\mathcal{P}$ and $\mathcal{C}$ are indicated by blue and red tubes in $(b_1, b_2, \theta)$ space. The singular support of $\mathcal{C}$ is also indicated; it is the $b_1$-$b_2$ projection of the wavefront set. Due to the additional dimension, the wavefront sets are more distinctly separated than the corresponding singular supports.

Living in this phase space is the wavefront set $WF(f)$; roughly, this is the set of position-orientation pairs at which $f$ is nonsmooth; for more details, see: [29, 7, 31].

Under the geometric separation model of Section 1, we have

$$WF(\mathcal{P}) = \mathrm{supp}(\mathcal{P}) \times \mathbf{P}^1,$$

since a point singularity is singular in all directions on its singular support, and is singular nowhere else; while

$$WF(\mathcal{C}) = \{(\tau(t), \theta(t)) : t \in [0, L(\tau)]\}$$

where $\tau(t)$ is a unit-speed parametrization of $\mathcal{C}$ and $\theta(t)$ is the normal direction to $\mathcal{C}$ at $\tau(t)$ regarded in $\mathbf{P}^1$.

It is convenient to think of the parameter space for microlocal analysis as a plane $\mathbf{R}^2$ of positions $b$ lying beneath a third dimension of orientations $\theta$. Then the wavefront set of a point singularity $\mathcal{P}$ concentrated on a single point $x_1$ is a vertical line segment $\{x_1\} \times \mathbf{P}^1$, corresponding to singular behavior in every direction at a given point, while the wavefront set of $\mathcal{C}$ is a more general curve in phase space. Even if $x_1$ meets $image(\tau)$, so the singular support of the point singularity and a curvilinear singularity overlap at $x_1$, they behave quite differently as wavefront sets in the full 3D parameter space, which gives us hope for separation.

Figure 2 illustrates the 3D phase space with pointlike and curvelike singularities superposed.

### 3.2 Support of Frame Elements

A radial wavelet $\psi_\lambda$, $\lambda = (a_0, b_0)$ is 'morally' supported in $x$-space in a spatial ball $B(a_0, b_0)$, defined by

$$B(a_0, b_0) = \{x : |x - b_0|/a \le 1\};$$

this statement should not however be taken too literally, as the radial wavelets we study are not of compact support. The more precise statement is that the wavelet decays rapidly in the variable $|x - b_0|/a$. The next statement uses the notation

$$\langle a \rangle = (1 + a^2)^{1/2}.$$

**Lemma 3.1** *For each $N = 1, 2, \ldots$ there is a constant $c_N$ so that*

$$|\psi_{a,b}(x)| \le c_N \cdot a^{-1} \cdot \langle |x - b|/a \rangle^{-N}, \qquad \forall a \in \mathbf{R}^+ \; \forall b, x \in \mathbf{R}^2.$$

An individual curvelet $\gamma_\eta$, $\eta = (a_0, b_0, \theta_0)$ is 'morally' supported in $x$-space inside an anisotropic spatial ellipse. To make this precise, let $D_{1/a}$ be the diagonal matrix $diag(1/a, 1/\sqrt{a})$ and $R_{-\theta}$ denote planar rotation by $-\theta$ radians. Let

$$P_{a,\theta} = D_{1/a} R_{-\theta}$$

denote the parabolic directional dilation operator which dilates much more strongly in the $\theta$ direction than in the orthogonal direction. For a vector $v \in \mathbf{R}^2$ define the norm

$$|v|_{a,\theta} = |P_{a,\theta}(v)|;$$

the unit ball in this norm is ellipsoidal, with minor axis pointing in direction $\theta$. A curvelet is morally supported in the ellipse $E(a_0, b_0, \theta_0)$ defined by

$$E(a_0, b_0, \theta_0) = \{x : |x - b_0|_{a_0,\theta_0} \le 1\},$$

Again the correct formal statement is that there is rapid decay in the variable $|x - b|_{a,\theta}$. The following is proved in [7, Lemma 1, page 168].

**Lemma 3.2** *For each $N = 1, 2, \ldots$ there is a constant $c_N$ so that*

$$|\gamma_{a,b,\theta}(x)| \le c_N \cdot a^{-3/4} \cdot \langle |x - b|_{a,\theta} \rangle^{-N}, \qquad \forall a \in \mathbf{R}^+ \; \forall \theta \in [0, \pi) \; \forall b, x \in \mathbf{R}^2. \tag{3.1}$$

In Figure 3, we visualize these support relationships.

### 3.3 Phase-Space Support of Frame Elements

We also study the location-orientation behavior of frame elements, i.e., the attribution of regions in an orientation/location domain as regions of significant activity in a distribution $f$. The wavefront set gives a qualitative way to do this; we use the continuous curvelet transform (CCT) to do so quantitatively. This transform is defined by

$$\Gamma_f(a, b, \theta) = \langle \gamma_{a,b,\theta}, f \rangle$$

and is indexed by triples $(a, b, \theta)$, where $a > 0$, $b \in \mathbf{R}^2$, and $\theta \in [0, \pi)$; this associates a function $f$ to a scale/location/direction domain. There is a natural measure on this domain:

$$m(da, db, d\theta) = a^{-3} \cdot da \cdot db \cdot d\theta.$$

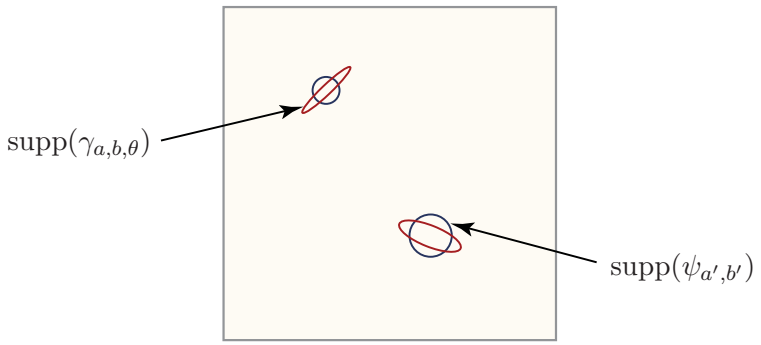


Figure 3: Effective supports of wavelets and curvelets at different scales. The shapes of those supports become increasingly distinct with increasingly fine scales.

Consider a high-pass function $f$, i.e., one whose Fourier transform of $f$ vanishes near the origin; $\hat{f}(\xi) = 0$, $|\xi| \leq \pi/a_0$, see [7]. The CCT offers a Parseval relationship for high-pass functions:

$$\int\int\int |\Gamma_f(a,b,\theta)|^2 m(da,db,d\theta) = \|f\|_2^2,$$

see [8]. Hence the energy in $f$ is distributed through the scale-location-direction domain by the curvelet transform offering a portrait of the function's significant activity.

Consider the transform of a radial wavelet:

$$\Gamma_{\psi_{a_0,b_0}}(a,b,\theta) = \langle \gamma_{a,b,\theta}, \psi_{a_0,b_0} \rangle, \qquad a > 0,\ b \in \mathbf{R}^2,\ \theta \in [0,\pi).$$

Because $\psi_{a_0,b_0}$ is radial, $\Gamma_{\psi_{a_0,b_0}}(a,b,\theta)$ is constant, independent of $\theta$, and decays rapidly in variables $|\log_2(a/a_0)|$ and $|b-b_0|/a_0$. It is morally localized to a cell of the form

$$W(a_0,b_0) = \{(a,b,\theta) : |\log_2(a/a_0)| < 1, \theta \in [0,\pi), |b-b_0|/a_0 \leq 1\};$$

we have the following formal statement, proved in Subsection 9.2.1 below:

**Lemma 3.3** *For each $N = 1,2,\ldots$, there is a constant $c_N$ so that*

$$|\langle \gamma_{a,b,\theta}, \psi_{a_0,b_0} \rangle| \leq c_N \cdot a^{1/4} \cdot 1_{\{|\log_2(a/a_0)|<3\}} \cdot \langle |b-b_0|_{a,\theta} \rangle^{-N}.$$

It implies the following for a scale-conditional phase portrait of a wavelet (i.e., we freeze the analysis scale $a$ at a specific value, and inspect $\Gamma_{\psi_{a_0,b_0}}(a,b,\theta)$ as a function of variables $b$ and $\theta$): when freezing $a = a_0$, we see that $\Gamma$ is 'morally' supported in a vertical tube above the point $b_0$; each horizontal cross-section is a ball of width $a_0$, i.e., $B(a_0,b_0)$.

Consider now the transform of a curvelet:

$$\Gamma_{\gamma_{a_0,b_0,\theta_0}}(a,b,\theta) = \langle \gamma_{a,b,\theta}, \gamma_{a_0,b_0,\theta_0} \rangle, \qquad a > 0,\ b \in \mathbf{R}^2,\ \theta \in [0,\pi).$$

This is 'morally' localized to a cell of the following form:

$$Q(a_0,b_0,\theta_0) = \{(a,b,\theta) : |\log_2(a/a_0)| < 1, |\theta-\theta_0| < \sqrt{a_0}, |b-b_0|_{a_0,\theta_0} \leq 1\};$$

the correct formal statement being [see (21) in [8]]

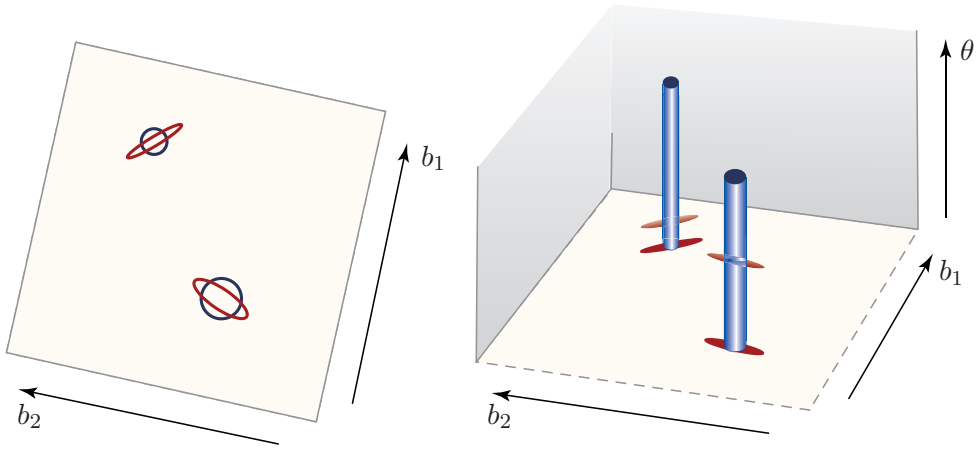


Figure 4: Left panel: effective supports of wavelets (blue) and curvelets (red); compare Figure 3. Right Panel: Phase space portrait – obtained by CCT – of the same wavelets and curvelets. Effective supports of curvelets are indicated in the $(b_1, b_2)$ plane. The portrait conveys the intuition that while the effective support of wavelets and curvelets may overlap substantially in the spatial domain, in the full phase space, the relative overlap is significantly reduced.

**Lemma 3.4** *For each $N = 1, 2, \ldots$, there is a constant $c_N$ so that*

$$|\langle \gamma_{a,b,\theta}, \gamma_{a_0,b_0,\theta_0} \rangle| \le c_N \cdot 1_{\{|\log_2(a/a_0)|<3\}} \cdot 1_{\{|\theta-\theta_0|<10\sqrt{a_0}\}} \cdot \langle |b - b_0|_{a_0,\theta_0} \rangle^{-N}.$$

(We remind the reader of our convention that, for two points $\theta, \theta' \in [0, \pi)$, $|\theta - \theta'|$ really means geodesic distance in $\mathbf{P}^1$.)

Thus each curvelet is supported in scale, location, direction in a set which effectively has a product structure, and is compactly supported in both scale and orientation. In a scale-conditional phase portrait, freezing $a = a_0$, we see a vaguely ellipsoidal structure with slice $\theta = \theta_0$ exhibiting an anisotropic footprint, like $E(a_0, b_0, \theta_0)$.

In Figure 4 we visualize the scale-conditional portraits of a wavelet and a curvelet. The intuition to be fostered from these figures is that curvelets don't interact very heavily with wavelets, because they have such different support in $(b, \theta)$, even when they have the same scale and location parameters, $a_0$ and $b_0$.

This support structure of curvelets is notable for other reasons. The support structure is implicit in the natural measure:

$$\begin{aligned} m(da, db, d\theta) &= a^{-3} \cdot da \cdot db \cdot d\theta \\ &= \frac{da}{a} \cdot \frac{db}{a^{3/2}} \cdot \frac{d\theta}{a^{1/2}}, \qquad 0 < a < a_0, b \in \mathbf{R}^2,\ \theta \in \mathbf{P}^1. \end{aligned}$$

The second expression has the following interpretation. We assign roughly unit measure to curvelet cells

$$m(Q(a, b, \theta)) \approx 1, \qquad 0 < a < a_0,\ b \in \mathbf{R}^2, \theta \in \mathbf{P}^1;$$

there are about $a^{-3/2}$ locations per unit volume at a fixed scale and orientation and about $a^{-1/2}$ orientations at a fixed scale and location.

The support structure is mimicked by the discretization of the curvelet tight frame, which morally breaks the $(a, b, \theta)$ domain into disjoint cells $Q(a_j, b_{j,k,\ell}, \theta_{j,\ell})$ and samples the continuous transform once per cell; for details see [8].

### 3.4 CCT of Singularities

The CCT can be used to analyze the singularities $\mathcal{P}$ and $\mathcal{C}$ directly. Let's impose regularity conditions on $\mathcal{C}$.

Let the Hausdorff pseudo-distance $d$ between $b$ and a curve $\tau$ be defined by

$$d(a, b, \theta; \tau) = \min\{|x - b|_{a,\theta} : x \in image(\tau)\}.$$

**Definition 3.1** *A finite-length planar curve $\tau$ will be called* parabolically regular*, if, for $N = 1, 2, \ldots$, there is a constant $c_N$ so that for $a \in (0, 1)$ and all $b, \theta$,*

$$\int \langle |\tau(t) - b|_{a,\theta} \rangle^{-N} dt \le c_N \cdot a^{1/2} \cdot \langle d(a, b, \theta; \tau) \rangle^{-N}. \tag{3.2}$$

Traditional nice curves, such as line segments, circles, etc. are parabolically regular, though we skip the demonstration.

**Lemma 3.5** *Let the singularity $\mathcal{C}$ be defined as in (1.2) by a parabolically regular curve $\tau$. Then, for each $N = 1, 2, \ldots$,*

$$|\Gamma_{\mathcal{C}}(a, b, \theta)| \le c_N \cdot a^{-1/4} \cdot \langle d((a, b, \theta); \tau) \rangle^{-N}.$$

In words, the CCT may be large at phase space points close to $\tau$, but elsewhere it is very small.

**Proof.** The definition of $\mathcal{C}$ as a linear functional on the Schwartz space gives

$$\Gamma_{\mathcal{C}}(a, b, \theta) = \int \gamma_{a,b,\theta}(\tau(t)) dt.$$

The estimates (3.1) and (3.2) give

$$\begin{aligned} |\Gamma_{\mathcal{C}}(a, b, \theta)| &\le \int c_N \cdot a^{-3/4} \cdot \langle |\tau(t) - b|_{a,\theta} \rangle^{-N} dt \\ &\le c_N \cdot a^{-3/4} \cdot C'_N \cdot a^{1/2} \cdot \langle d(a, b, \theta; \tau) \rangle^{-N} \\ &= C''_N \cdot a^{-1/4} \cdot \langle d(a, b, \theta; \tau) \rangle^{-N}. \square \end{aligned}$$

### 3.5 Heuristics

We are now in a position to give a heuristic explanation why the strategy announced in Section 2.2 is likely to work. In effect, there are very instructive analogies between the calculations needed to implement that strategy and the behavior of certain 'tubes' in phase space. The reader will have noticed that the curvelet parameter $(a, b, \theta)$ and the wavefront set parameter $(b, \theta)$ differ only by the latter's provision of a scale. Hence there is some

analogy between the scale-conditional portrait by CCT and the wavefront set – both provide measures of the activity of an object, indexed by location and orientation.

In effect, the scale-conditional portrait by CCT is a "thickened-out" version of the wavefront set. A point singularity has a wavefront set which is a vertical line in phase space, and a scale-conditional portrait which is localized near a thin vertical tube. A curvilinear singularity has a wavefront set which is a curve in phase space, while Lemma 3.5 says that, morally, the curvelet transform of the object $\mathcal{C}$ 'lives' near a tube. The tube in question has thickness $\approx a$. Look at the scale-conditional portrait, and define the tube

$$T(a_0) = \bigcup_{b,\theta} (E(a_0, b, \theta) \times \{\theta\}),$$

where the union is over $b$, $\theta$ satisfying

$$d((a_0, b, \theta); \tau) \le 1.$$

As $a_0 \to 0$, this tube shrinks down to a curve $\tilde{\tau}$ in phase space $\mathbf{R}^2 \times \mathbf{P}^1$ defined by

$$\tilde{\tau}(t) = (\tau(t), \theta(t)),$$

where $\theta(t)$ is the orientation of the normal to $\tau(t)$. In short, in the sense of set convergence

$$T(a_0) \to WF(\mathcal{C}), \qquad a_0 \to 0.$$

Thus, the wavefront set and the curvelet transform both signal that the activity in location-orientation space is concentrated near $image(\tilde{\tau})$.

More is true. A wavelet has a scale-conditional portrait which is a thin vertical tube – similar to the phase portrait of a point singularity – while a curvelet has a scale-conditional portrait which is a tube surrounding a little 'piece of a curve' in phase space, i.e., it morally has a position and orientation.

This visual analogy suggests that curvelets are incoherent to wavelets – because of the low overlap in phase space. Indeed, from Parseval,

$$\int\int\int \Gamma^{\star}_{\psi_{a_0,b_0}}(a, b, \theta) \Gamma_{\gamma_{a_1,b_1,\theta_1}}(a, b, \theta) m(da, db, d\theta) = \langle \psi_{a_0,b_0}, \gamma_{a_1,b_1,\theta_1} \rangle,$$

and so the low overlap between the two phase portraits indeed will cause relatively low singleton coherence (2.2); indeed the tubelet associated to a given curvelet and the tube associated to a given wavelet visibly have relatively small overlap in the scale-conditional phase portrait; for example, if we compare the overlap of effective supports in phase space to the overlap of effective supports in the spatial domain, we see that the fractional overlap is dramatically smaller at fine scales in the phase space portrait than it is in the spatial domain portrait.

However, the singleton coherence is not sufficiently small to be powerful in the present setting. Instead, this paper develops cluster coherence. The visual analogy presented in Figure 5 suggests how to bound the cluster coherence and suggests that the proof strategy of Section 2.2 will succeed. To understand that analogy, let's study Figure 5. If we let $\mathcal{S}_1$ denote the set of significant wavelet coefficients in the radial wavelet transform of $\mathcal{P}$ at scale $a_0$, and $\mathcal{S}_2$ denote the set of significant curvelet coefficients in the curvelet transform of $\mathcal{C}$ at scale $a_0$, we believe the reader will be easily able to motivate the following assertions on the basis of Figures 3-5:

- Wavelets in $\mathcal{S}_1$ are associated to vertical tubes clustering around the point singularities in $\mathcal{P}$;
- Curvelets in $\mathcal{S}_2$ are associated with tubes clustering around the curvilinear phase portrait of $\mathcal{C}$;
- No single wavelet's phase portrait overlaps much with the cluster of curvelet phase portraits of $\mathcal{S}_2$;
- No single curvelet phase portrait overlaps with the cluster of wavelets in $\mathcal{S}_1$.

Let's outline a pseudo-calculation inspired by these visual observations. First, we consider a pseudo-calculation of the cluster coherence,

$$\mu_c(\mathcal{S}_{1,j}, \textsc{Wavelet scale } j; \textsc{Curvelet scale } j) = \sup_{\eta} \sum_{\lambda \in \mathcal{S}_{1,j}} |\langle \gamma_\eta, \psi_\lambda \rangle|.$$

With $(\psi_i)_i$ an enumeration of the significant wavelets in the expansion at scale $a_j = 2^{-j}$,

$$\begin{aligned} \sum_i |\langle \gamma_\eta, \psi_i \rangle| &= \sum_i \left| \int\int\int \Gamma^*_{\gamma_\eta} \Gamma_{\psi_i} dm \right| \\ &\le \sum_i \int\int\int |\Gamma_{\gamma_\eta}|(a,b,\theta) |\Gamma_{\psi_i}|(a,b,\theta) dm \\ &\le \int\int\int |\Gamma_{\gamma_\eta}|(a,b,\theta) \sum_i |\Gamma_{\psi_i}|(a,b,\theta) dm. \end{aligned}$$

Now use the bounds on $|\Gamma_{\psi_i}|(a,b,\theta)$ given above in Lemma 3.3, and deploy the slogan that *only a bounded number of significant wavelets at any given scale interact strongly with any specific phase space point*; we have

$$\sum_i |\Gamma_{\psi_i}|(a,b,\theta) \lesssim C_1 \cdot a_j^{1/4},$$

where the sum is over the significant wavelets at scale $a_j$ and $C_1$ does not depend on $j$, and the symbol $\lesssim$ indicates an inequality motivated heuristically – in this case by the preceding italicized slogan. We also observe that Lemma 3.4 implies that the integral over phase space of a curvelet phase portrait obeys $\int\int\int |\Gamma_{\gamma_\eta}|(a,b,\theta) dm \le C_2$. Combining this with the previous displays, we pseudo-conclude that

$$\mu_c(\mathcal{S}_1, \textsc{Wavelet scale } j; \textsc{Curvelet scale } j) \to 0, \qquad j \to \infty.$$

Next, we consider a pseudo-calculation of the cluster coherence

$$\mu_c(\mathcal{S}_2, \textsc{Curvelet scale } j; \textsc{Wavelet scale } j) = \sup_{\lambda} \sum_{\eta \in \mathcal{S}_{2,j}} |\langle \gamma_i, \psi_\lambda \rangle|.$$

With $(\gamma_i)_i$ an enumeration of the significant curvelets in the expansion at scale $a_j = 2^{-j}$, then for a fixed wavelet index $\lambda$ we have

$$\begin{aligned}\sum_i |\langle \gamma_i, \psi_\lambda \rangle| &= \sum_i \left| \int\int\int \Gamma^*_{\gamma_i} \Gamma_{\psi_\lambda} dm \right| \\ &\le \sum_i \int\int\int |\Gamma_{\gamma_i}|(a,b,\theta)|\Gamma_{\psi_\lambda}|(a,b,\theta) dm \\ &\le \int\int\int |\Gamma_{\psi_\lambda}|(a,b,\theta) \sum_i |\Gamma_{\gamma_i}|(a,b,\theta) dm.\end{aligned}$$

Now use the bounds on $|\Gamma_{\gamma_i}|(a,b,\theta)$ given above in Lemma 3.4, and apply the slogan that *only a bounded number of significant curvelets at any given scale interact strongly with a given phase space point.* We have

$$\sum_i |\Gamma_{\gamma_i}|(a,b,\theta) \lesssim C_1,$$

where the sum is over the significant curvelets at scale $a_j$ and $C_1$ is a constant. We also observe that Lemma 3.3 implies that the integral over phase space of a wavelet phase portrait obeys $\int\int\int |\Gamma_{\psi_\lambda}|(a,b,\theta) dm \le C_2 \cdot a_0^{1/4}$, where $a_0 = a_0(\lambda)$. Combining this with the previous displays, we pseudo-conclude that

$$\mu_c(\mathcal{S}_2, \textsc{Curvelet scale } j; \textsc{Wavelet scale } j) \to 0, \qquad j \to \infty.$$

We now turn to the approximation tasks posed by the strategy announced in Section 2.2. To pseudo-bound $\delta_{j,1}$, we fix $\varepsilon > 0$ and define the tube in phase space $\mathcal{T}_{j,1}$ consisting of all scale/location pairs $(a,b)$ where the bound provided in Lemma 4.1 permits coefficients larger than $a_j^{1-\varepsilon}$. Also, we let $i_j$ denote the index in the wavelet enumeration beyond which such potentially significant coefficients can no longer arise. We can heuristically approximate a sum of wavelet coefficients with an integral over the phase space region covered by the union of their phase space supports; then we have

$$\begin{aligned}\delta_{j,1} &= \sum_{i > i_j} |\langle \psi_i, \mathcal{P}_j \rangle| \\ &\lesssim c_N \cdot a_j^{-1/2} \cdot \int\int\int_{\mathcal{T}^c_{j,1}} \langle b/a_j \rangle^{-N} dm.\end{aligned}$$

For sufficiently large $N$, the integrand has powerful decay; for large $j$, the width of the tube $\mathcal{T}_{j,1}$ is significantly wider than the decay scale $a_j$; and so the integral becomes negligible for large $j$, i.e., we pseudo-conclude that $\delta_{j,1} \to 0$. To pseudo-bound $\delta_{j,2}$ we fix $\varepsilon > 0$ and define the tube in phase space $\mathcal{T}_{j,2}$ consisting of all triples $(a,b,\theta)$ where the bound provided in Lemma 3.5 permits coefficients larger than $a_j^{1-\varepsilon}$. Also we let $i_j$ denote the index in the curvelet enumeration beyond which such potentially significant coefficients can no longer arise; again heuristically identifying a sum with a phase-space integral we have

$$\begin{aligned}\delta_{j,2} &= \sum_{i > i_j} |\langle \gamma_i, \mathcal{C}_j \rangle| \\ &\lesssim c_N \cdot \int\int\int_{\mathcal{T}^c_{j,2}} a^{-1/4} \cdot \langle d(a_j, b, \theta; \tau) \rangle^{-N} dm.\end{aligned}$$

For large $N$ the integrand decays strongly; for large $j$ the width of the tube is significantly wider than the decay scale $a_j$; and so again the integral becomes negligible for large $j$, i.e., we pseudo-conclude that $\delta_{j,2} \to 0$.

In short, phase space diagrams, some elementary estimates motivated by tube overlaps, and some cardinality 'slogans' combine to show plausibility of the strategy announced in Section 2.2. In the sections to come, we rigorously carry out that strategy. While the details are much more delicate than this plausibility argument would suggest, the architecture of our full demonstration remains faithful to the geometric viewpoint.

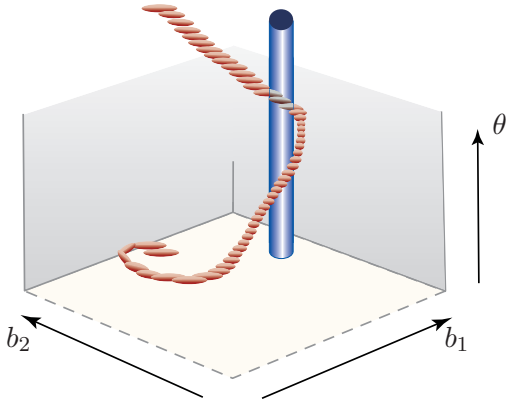


Figure 5: Phase space portrait of a cluster of curvelets and one single wavelet. Note the visually small overlap of these two geometrical objects. This suggests by analogy that each single wavelet is not coherent with anything built from such a cluster of curvelets.

# 4 The Cluster $\mathcal{S}_{1,j}$ and its estimates

In this section, we define the cluster of wavelet coefficients $\mathcal{S}_{1,j}$ of the filtered point singularity $F_j \star \mathcal{P}$, and estimate relative sparsity as well as cluster coherence using this cluster. We intend to show that with this definition of cluster set,

$$\delta_{1,j} = o(\|f_j\|_2), \qquad j \to \infty, \tag{4.1}$$

and

$$\mu_c(\mathcal{S}_{1,j}, \{\psi_\lambda\}; \{\gamma_\eta\}) \to 0, \qquad j \to \infty. \tag{4.2}$$

As explained in Section 2.2, this gives the needed part of Theorem 1.1 having to do with $\mathcal{S}_{1,j}$.

WLOG we can assume that

$$\mathcal{P} = |x|^{-3/2}.$$

The result for the more general $\mathcal{P}$ of (1.1) follows easily by combining translation invariance with finitely many uses of the triangle inequality. We also, from now on, fix some

$$\varepsilon \in (0, 1/32).$$

Our first lemma is used frequently in what follows, and is crucial for our definition of the cluster of wavelet coefficients. For the proof, see Subsection 9.3.1.

**Lemma 4.1** *For each $N = 1, 2, \ldots$, there is a constant $c_N$ so that*

$$|\langle \psi_{a_{j'},b}, \mathcal{P}_j \rangle| \le c_N \cdot 2^{j/2} \cdot 1_{\{|j-j'|\le 1\}} \cdot \langle |b/a_j| \rangle^{-N}, \qquad \forall j, j' \in \mathbf{Z}_+, \ \forall b \in \mathbf{R}^2.$$

In line with the heuristics of the previous section, we think of our estimate as describing relative overlaps of tubes in phase space. However, in the particular case of wavelets, there is no directional selectivity, so all that matters is the projection of phase space onto the spatial domain. We measure spatial distances with

$$d_2(x, A) = \min_{a \in A} \|x - a\|_2, \quad x \in \mathbf{R}^2, \ A \subseteq \mathbf{R}^2,$$

the Euclidean distance between a point $x$ and a set $A$.

Of course, since we are dealing with frames, ultimately we have to consider discrete indices. To support geometric intuition, most of our arguments will be in the continuum setting, restrictions to discrete sampling grids being delayed as late in each argument as possible.

Morally, the points in phase space associated with significant wavelet coefficients are contained in a tube around $WF(\mathcal{P})$ in phase space. This neighborhood of $WF(\mathcal{P})$ can be explicitly defined by

$$\mathcal{N}_1^{PS}(a) = \{b \in \mathbf{R}^2 : d_2(b, \{0\}) \le D_1(a)\} \times [0, \pi),$$

where

$$D_1(a) = a^{(1-\varepsilon)}.$$

The shape of the tube reflects the isotropic behavior of $WF(\mathcal{P})$. For an illustration of $\mathcal{N}_1^{PS}(a)$, we refer to Figure 6.

We define the cluster of wavelet coefficients around the point-singularity by intersecting the tube $\mathcal{N}_1^{PS}(a)$ with the wavelet lattice, i.e.,

$$\mathcal{S}_{1,j} = \{(j,k) \in \Delta_j^{\pm 1} : \{b_{j,k}\} \times [0, \pi) \in \mathcal{N}_1^{PS}(a_j)\}.$$

The remainder of the section establishes (4.1)-(4.2) for this definition of cluster.

## 4.1 Size of $f_j$

**Lemma 4.2** *For some $c > 0$,*

$$\|f_j\|_2 \ge c 2^{j/2}, \qquad j \to \infty.$$

**Proof.** Apply (1.3) and (1.4):

$$\|f_j\|_2^2 \;=\; \int_{\mathbf{R}^2} W^2(|\xi|/2^j) |\hat{f}(\xi)|^2 d\xi \ge c \cdot \int_{\mathcal{A}_j} |\xi|^{-1} d\xi \ge c \cdot 2^j. \qquad \square$$

### 4.2 $\mathcal{S}_{1,j}$ offers low approximation error

Now we are ready to state and prove the approximation error of $\mathcal{S}_{1,j}$.

**Lemma 4.3**

$$\delta_{1,j} = \sum_{\lambda \in \mathcal{S}_{1,j}^c} |\langle \psi_\lambda, \mathcal{P}_j \rangle| = o(\|f_j\|_2), \qquad j \to \infty.$$

**Proof.** Due to the specific filtering we use,

$$\delta_{1,j} = \sum_{\lambda \in \mathcal{S}_{1,j}^c} |\langle \psi_\lambda, \mathcal{P}_j \rangle| = \sum_{j'=j-1}^{j+1} \sum_{\{k:\{b_{j',k}\}\times[0,\pi) \notin \mathcal{N}_1^{PS}(a_{j'})\}} |\langle \psi_{(j',k)}, \mathcal{P}_j \rangle|.$$

Applying Lemma 4.1, and picking $N$ so large that $(N-1)\varepsilon > 1/2$

$$\delta_{1,j} \le c \cdot \sum_{|k|>2^{j\varepsilon}} c_N \cdot 2^{j/2} \cdot \langle |k| \rangle^{-N} = o(1), \qquad j > j_0.$$

So the lemma is proved. □

### 4.3 $\mathcal{S}_{1,j}$ offers low cluster coherence

**Lemma 4.4**

$$\mu_c(\mathcal{S}_{1,j}, \{\psi_\lambda\}; \{\gamma_\eta\}) \to 0, \qquad j \to \infty.$$

**Proof.** By Lemma 3.3, if $\lambda \in \Lambda_j^{\pm 1}$,

$$|\langle \psi_\lambda, \gamma_\eta \rangle| \le c \cdot 2^{-j/4}, \qquad \forall \eta.$$

The definition of $\mathcal{S}_{1,j}$ via $\mathcal{N}_1^{PS}(a_j)$ implies that

$$\#(\mathcal{S}_{1,j}) \le c \cdot 2^{2j\varepsilon}, \qquad \text{for sufficiently large } j.$$

We conclude from $\varepsilon < 1/8$ that

$$\sup_\eta \sum_{\lambda \in \mathcal{S}_{1,j}} |\langle \psi_\lambda, \gamma_\eta \rangle| \le c \cdot 2^{-j(1/4-2\varepsilon)} \to 0, \qquad j \to \infty. \qquad \square$$

## 5 Sparse Expansion of a Linear Singularity

After the relative ease with which we obtained concentration estimates (4.1)-(4.2) for the cluster of significant wavelet coefficients, we must now brace ourselves for the considerably harder challenge posed by the analogous estimates for the cluster of significant curvelet coefficients. This extra work seems, at least to us, much more rewarding, as it involves a full-blown use of phase space geometry.

In this section, we develop essential infrastructure for the analysis to come, documenting the sparsity of curvelet coefficients of a special linear singularity. Let $w_2 : \mathbf{R} \mapsto [0,1]$ be a

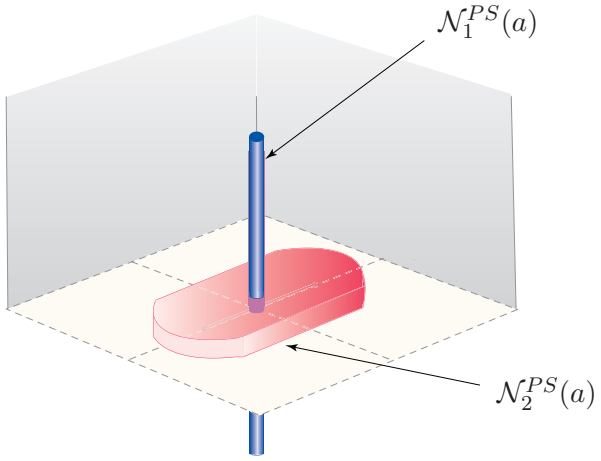


Figure 6: The tubes $\mathcal{N}_1^{PS}(a)$ and $\mathcal{N}_2^{PS}(a)$ in phase space. The clusters of significant coefficients correspond roughly to phase-space support regions overlapping these tubes.

smooth function to be specified later (cf. Subsection 6.2), supported in $[-1,1]$, and define the very special distribution $w\mathcal{L}$ supported on a line segment $\{0\} \times [-\rho, \rho]$ by

$$w\mathcal{L} = w_2(x_2/\rho) \cdot \delta_0(x_1).$$

Then we can write

$$\widehat{w\mathcal{L}} = \hat{w} \star \hat{\mathcal{L}},$$

where

$$\hat{w} = \hat{w}_2(\rho\xi_2) \cdot \rho \cdot \delta_0(\xi_1) \quad \text{and} \quad \hat{\mathcal{L}} = \delta_0(\xi_2).$$

Thus the action of $w\mathcal{L}$ on a continuous function $f$ is given by

$$2\pi\langle w\mathcal{L}, f\rangle = \langle \mathcal{L}, \hat{w} \star \hat{f}\rangle = \int (\hat{w} \star \hat{f})(\xi_1, 0) d\xi_1. \tag{5.1}$$

Conceptually, $w\mathcal{L}$ is a straight curve fragment; our analysis of $\mathcal{C}$ in Section 6 will reduce to the study of this case.

Define a tube in phase space, in which the significant curvelet coefficients will be located. This will now be a neighborhood of $WF(w\mathcal{L})$, defined by

$$\mathcal{N}_2^{PS}(a) = \{b \in \mathbf{R}^2 : d_2(b, \{0\} \times [-2\rho, 2\rho]) \le D_2(a)\} \times [0, \sqrt{a}],$$

where

$$D_2(a) = a^{(1-\varepsilon)}.$$

For an illustration of $\mathcal{N}_2^{PS}(a)$, and its relation to $\mathcal{N}_1^{PS}(a)$, we refer to Figure 6. The actual definition of the cluster of curvelet coefficients is much more involved. In Lemma 5.4, we will introduce a first set which helps to determine its location.

Several bounds will control the curvelet coefficients of a linear singularity. Lemma 3.5 gives

$$|\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| \le c \cdot 2^{j/4}, \qquad \forall a, b, \theta; \tag{5.2}$$

in fact that lemma even gives a decay estimate, as the microlocation $(b, \theta)$ moves away from $(\{0\} \times [-\rho, \rho]) \times \{0\}$. In the situations where we would use that decay estimate, the next lemma is more convenient.

**Lemma 5.1** *Suppose that $\theta \in [0, \sqrt{a}]$, and set*

$$\tau := \cos\theta \sin\theta (a^{-1} - a^{-2}), \quad d_1^2 := b_1^2(\sigma_2^2 - \sigma_1^{-2}\tau),$$

*and*

$$d_2^2 := \begin{cases} \min_{\pm} \left((\pm\rho - b_2)\sigma_1 + \sigma_1^{-1} b_1 \tau\right)^2 & : \quad b_2 - \sigma_1^{-2} b_1 \tau \notin [-\rho, \rho], \\ 0 & : \quad b_2 - \sigma_1^{-2} b_1 \tau \in [-\rho, \rho], \end{cases}$$

*where*

$$\sigma_1 = (a^{-2}\sin^2\theta + a^{-1}\cos^2\theta)^{1/2} \quad \textit{and} \quad \sigma_2 = (a^{-1}\sin^2\theta + a^{-2}\cos^2\theta)^{1/2}.$$

*Then, for $N = 1, 2, \ldots$,*

$$|\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| \le c_N \cdot a^{-3/4} \cdot \sigma_1^{-1} \cdot \langle d_1\rangle^{-1} \cdot \langle |(d_1, \sigma_1 d_2)|\rangle^{2-N}.$$

In some cases, spatial decay alone is insufficient and we also need to exploit directional localization; for such cases we employ the following lemma.

**Lemma 5.2** *Suppose that $\theta \in (\sqrt{a}, \pi)$. Then, for $L, M = 0, 1, 2, \ldots$,*

$$\begin{aligned} |\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| \quad \le \quad & c_{L,M} \cdot a^{-1/4} \cdot |\cos\theta| \cdot e^{-\rho \frac{|\sin\theta|}{2a}} \cdot \langle |b_1|\rangle^{-L} \cdot (a^{1/2}|\sin\theta| + a|\cos\theta|)^L \\ & \cdot \langle |b_2|\rangle^{-M} \cdot (\rho + a^{1/2}|\cos\theta| + a|\sin\theta|)^M. \end{aligned}$$

Both previous lemmas will be proved in Subsections 9.4.1 and 9.4.2, respectively. Together, they imply that the curvelet frame coefficients of $w\mathcal{L}$ are sparse. Indeed, in the directional panels where $\theta$ is close to 0, we have about $2^{j/2}\rho$ significantly nonzero coefficients, which are bounded by $c2^{j/4}$, while in the directional panels where $\theta$ is far from zero, we have few significantly nonzero coefficients. Formally,

**Lemma 5.3** *Let $\alpha_j = (\langle w\mathcal{L}_j, \gamma_\eta\rangle)_\eta$ denote the curvelet frame coefficients of $w\mathcal{L}_j = F_j \star w\mathcal{L}$. For each $p > 0$, there is $c_p > 0$ so that,*

$$\|\alpha_j\|_p \le c_p \cdot 2^{j((1/2+\varepsilon)/p + 1/4)}, \qquad j > j_0.$$

This will be proved in Section 5.1. The next result, making precise the location of the significant coefficients, is proved in Section 5.2.

**Lemma 5.4** *Put*

$$\tilde{\mathcal{S}}_j = \{(j, k, \ell) \in \Delta_j^{\pm 1} : (b_{j,k,\ell}, \theta_{j,\ell}) \in \mathcal{N}_2^{PS}(a_j)\}.$$

*Then*

$$\|\alpha_j 1_{(\tilde{\mathcal{S}}_j)^c}\|_1 = O(1), \qquad j \to \infty.$$

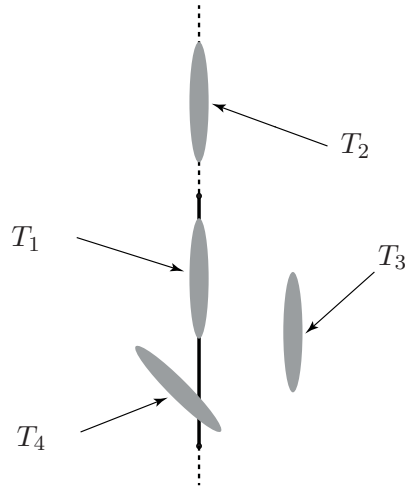


Figure 7: Curvelets associated with the cases $T_1$–$T_4$. The line segment is the support of $w\mathcal{L}_j$. The dotted line is the affine extension of that segment.

### 5.1 Proof of Lemma 5.3

We first observe that the full curvelet coefficient vector $\alpha_j$ is simply the extension of $(\langle w\mathcal{L}_j, \gamma_\eta\rangle)_{\eta\in\Delta_j^{\pm 1}}$ to scales away from $\{j-1, j, j+1\}$ by zero filling. Also WLOG we can assume that $\eta \in \Delta_j$, since the terms related to the scales $j-1$ and $j+1$ only change the constant factor of the final estimate independent on $j$.

Define the following four regions in phase space:

$$\begin{aligned}
\mathcal{N}_2^{PS}(a) &= \{b \in \mathbf{R}^2 : d_2(b, \{0\}\times[-2\rho, 2\rho]) \le D_2(a)\} \times [0, \sqrt{a}],\\
\mathcal{N}_3^{PS}(a) &= (\{b \in \mathbf{R}^2 : d_2(b, \{0\}\times\mathbf{R}) \le D_2(a)\} \times [0, \sqrt{a}]) \setminus \mathcal{N}_2^{PS}(a),\\
\mathcal{N}_4^{PS}(a) &= (\mathbf{R}^2 \times [0, \sqrt{a}]) \setminus (\mathcal{N}_2^{PS}(a) \cup \mathcal{N}_3^{PS}(a)),\\
\mathcal{N}_5^{PS}(a) &= \mathbf{R}^2 \times (\sqrt{a}, \pi).
\end{aligned}$$

We now split the norm $\|\alpha_j\|_p$ according to each curvelet coefficient's microlocation. For a phase space set $\mathcal{N}$ write $\eta \sim \mathcal{N}$, meaning the set $\{\eta : \eta \in \Delta_j \text{ and } (b_{j,k,\ell}, \theta_{j,\ell}) \in \mathcal{N}\}$. We have the following decomposition:

$$\begin{aligned}
\|\alpha_j\|_p^p &= \sum_{\eta\sim\mathcal{N}_2^{PS}(a_j)} |\langle w\mathcal{L}_j, \gamma_\eta\rangle|^p + \sum_{\eta\sim\mathcal{N}_3^{PS}(a_j)} |\langle w\mathcal{L}_j, \gamma_\eta\rangle|^p\\
&\quad + \sum_{\eta\sim\mathcal{N}_4^{PS}(a_j)} |\langle w\mathcal{L}_j, \gamma_\eta\rangle|^p + \sum_{\eta\sim\mathcal{N}_5^{PS}(a_j)} |\langle w\mathcal{L}_j, \gamma_\eta\rangle|^p. \qquad (5.3)
\end{aligned}$$

We have the following approximate equivalences:

$$\begin{aligned}
\eta \sim \mathcal{N}_2^{PS}(a_j) &\approx \{(j, (k_1, k_2), 0) : |k_1| \le 2^{j\varepsilon}, k_2 \in [-2\rho/\sqrt{a}, 2\rho/\sqrt{a}]\},\\
\eta \sim \mathcal{N}_3^{PS}(a_j) &\approx \{(j, (k_1, k_2), 0) : |k_1| \le 2^{j\varepsilon}, k_2 \notin [-2\rho/\sqrt{a}, 2\rho/\sqrt{a}]\},\\
\eta \sim \mathcal{N}_4^{PS}(a_j) &\approx \{(j, k, 0) : k \in \mathbf{Z}^2, |k_1| \ge 2^{j\varepsilon}\},\\
\eta \sim \mathcal{N}_5^{PS}(a_j) &\approx \{(j, k, \ell) : \ell = 1, ..., a_j^{-1/2} - 1\}.
\end{aligned}$$

Thus, in place of the original continuum-domain splitting (5.3), we consider instead the 'discrete-domain' splitting

$$\begin{aligned}\|\alpha_j\|_p^p &= \sum_{|k_1|\le 2^{j\varepsilon}} \sum_{|k_2|\le 2\rho/\sqrt{a}} |\langle w\mathcal{L}_j, \gamma_{j,(k_1,k_2),0}\rangle|^p + \sum_{|k_1|\le 2^{j\varepsilon}} \sum_{|k_2|>2\rho/\sqrt{a}} |\langle w\mathcal{L}_j, \gamma_{j,(k_1,k_2),0}\rangle|^p \\ &\quad + \sum_{|k_1|>2^{j\varepsilon}} |\langle w\mathcal{L}_j, \gamma_{j,k,0}\rangle|^p + \sum_{\ell=1}^{a_j^{-1/2}-1} \sum_{k\in\mathbf{Z}^2} |\langle w\mathcal{L}_j, \gamma_{j,k,\ell}\rangle|^p \\ &= T_1 + T_2 + T_3 + T_4. \end{aligned} \tag{5.4}$$

These terms correspond, respectively, to nearly vertical curvelets lying on the line segment singularity ($T_1$), nearly vertical curvelets centered elsewhere on the line containing the line segment ($T_2$), nearly vertical curvelets centered elsewhere ($T_3$), and all other curvelets ($T_4$), as illustrated in Figure 7.

To estimate $T_1$, use (5.2)

$$\sum_{k_2=-2\rho/\sqrt{a}_j}^{2\rho/\sqrt{a}_j} |\langle w\mathcal{L}_j, \gamma_{j,(0,k_2),0}\rangle|^p \le c\cdot 2^{j\frac{p}{4}} \cdot \sum_{k_2=-\rho/\sqrt{a}_j}^{\rho/\sqrt{a}_j} 1 \le c\cdot a_j^{-1/2+p/4}, \tag{5.5}$$

so $T_1 \le c\cdot 2^{j\varepsilon} a_j^{-1/2-p/4} = c2^{j(1/2+p/4+\varepsilon)}$. Here and below, when we write a sum taken over integers with non-integer bounds, we implicitly mean that the sum extends over all integers between the bounds.

To derive estimates for $T_2$–$T_3$, we first transfer the estimates for $\theta = 0$ derived in Lemma 5.1 (in terms of continuum parameters) to statements about $|\langle w\mathcal{L}_j, \gamma_{j,k,\ell}\rangle|$, in terms of the discrete lattice parameters.

**Lemma 5.5** *For $N = 1, 2, \ldots$,*

$$|\langle w\mathcal{L}_j, \gamma_{j,k,0}\rangle| \le c_N \cdot a_j^{-1/4} \cdot \langle |k_1|\rangle^{-1} \cdot \langle a_j^{-1}[|a_j k_1|^2 + \min_{\pm} |a_j^{1/2} k_2 \pm \rho|^2]^{1/2}\rangle^{2-N}.$$

**Proof.** This follows directly from the 'in particular'-part of Lemma 5.1 and the relation between continuous coefficients and lattice parameters given by (1.5). □

To estimate $T_2$, let $\mathcal{K}_2 = \{k_2 \in \mathbf{Z} : k_2 > 2\rho/\sqrt{a_j}\}$. Then

$$\begin{aligned}&\sum_{|k_2|>2\rho/\sqrt{a}_j} |\langle w\mathcal{L}_j, \gamma_{j,(0,k_2),0}\rangle|^p \\ &\le c_{N,p}\cdot a_j^{-p/4} \sum_{\mathcal{K}_2} \langle a_j^{-1} \min_{\pm} |a_j^{1/2} k_2 \pm \rho|\rangle^{(2-N)p} \\ &\le c_{N,p}\cdot a_j^{-p/4} \sum_{\mathcal{K}_2} \langle |a_j^{-1/2} k_2 - a_j^{-1}\rho|\rangle^{(2-N)p}. \end{aligned}$$

For $(N-2)p > 1$, we have

$$\int_{2\rho/\sqrt{a}_j}^{\infty} \langle |a_j^{-1/2} x - a_j^{-1}\rho|\rangle^{(2-N)p} dx = a_j^{1/2} \int_{\rho/a_j}^{\infty} \langle |y|\rangle^{(2-N)p} dy \le c_{N,p}\cdot a_j^{1/2+((N-2)p-1)}.$$

This estimate concerns $k_1 = 0$. For the other cases with $|k_1| < 2^{j\varepsilon}$ we use this same estimate, getting

$$T_2 \le c_{N,p} \cdot 2^{j(-1/2+\varepsilon)}. \tag{5.6}$$

To estimate $T_3$ let $\mathcal{K}_3 := \{k \in \mathbf{Z}^2, |k_1| > 2^{j\varepsilon}\}$ and choose $N$ so that $\varepsilon \cdot (N-2)p > 3\frac{1}{4}$. Then

$$\sum_{k\in\mathcal{K}_3} |\langle w\mathcal{L}_j, \gamma_{j,k,0}\rangle|^p \le c_N \cdot a_j^{-p/4} \sum_{k\in\mathcal{K}_3} \langle [|k_1|^2 + \min_{\pm} |a_j^{-1/2}k_2 \pm a_j^{-1}\rho|^2]^{1/2}\rangle^{(2-N)p}.$$

Partition the set $\mathcal{K}_3 = \mathcal{K}_3^0 \cup \mathcal{K}_3^1$, where $\mathcal{K}_3^0 = \{|k_2| < 2^j\rho\} \cap \mathcal{K}_3$. The sum over $\mathcal{K}_3^0$ involves sites where $k_2$ may as well be zero; it is not asymptotically larger than the LHS in this display:

$$4 \cdot \int_0^{2^j\rho} \int_{2^{j\varepsilon}}^{\infty} \langle |x_1|\rangle^{(2-N)p} dx_1 \quad = \quad 2^{j+2}\rho \cdot \int_{2^{j\varepsilon}}^{\infty} \langle |x_1|\rangle^{(2-N)p} dx_1 \le c \cdot 2^{j(1-\varepsilon\cdot((N-2)p-1))}.$$

Since $\varepsilon\cdot(N-2)p > 5/4+\varepsilon$, this last term is $O(2^{-j/4})$. The sum over $\mathcal{K}_3^1$ is not asymptotically larger than the LHS of the next display; the RHS uses Lemma 9.3:

$$\int_{2^j\rho}^{\infty} \int_{2^{j\varepsilon}}^{\infty} \langle |(x_1,x_2)|\rangle^{(2-N)p} dx_1 dx_2 \le c\langle |(2^{j\varepsilon}, 2^j\rho)|\rangle^{(2-N)p+2} \le 2^{j(1-\varepsilon\cdot((N-2)p-2))}$$

Since $\varepsilon \cdot (N-2)p > 3\frac{1}{4}$, this last term is $O(2^{-j/4})$. We conclude that

$$\sum_{k\in\mathcal{K}_3} |\langle w\mathcal{L}_j, \gamma_{j,k,0}\rangle|^p \le c_{N,p} \cdot a_j^{(1-p)/4}. \tag{5.7}$$

Before estimating $T_4$, we translate Lemma 5.2 into a simple form involving discrete curvelet parameters.

**Lemma 5.6** *Let*

$$\begin{aligned} T_{j,k,\ell} \quad = \quad & \langle |a_j k_1 \cos\theta_{j,\ell} - a_j^{1/2} k_2 \sin\theta_{j,\ell}|\rangle^{-1} \cdot (a_j^{1/2} |\sin\theta_{j,\ell}| + a_j |\cos\theta_{j,\ell}|) \\ & \cdot \langle |a_j k_1 \sin\theta_{j,\ell} + a_j^{1/2} k_2 \cos\theta_{j,\ell}|\rangle^{-1}. \end{aligned}$$

*There exist constants $c_N$ so that, for $j, k$, $N = 1, 2, \ldots$,*

$$|\langle w\mathcal{L}_j, \gamma_{j,k,\ell}\rangle| \le c_N \cdot a_j^{-1/4} \cdot e^{-\rho\frac{|\sin\theta_{j,\ell}|}{2a_j}} \cdot T_{j,k,\ell}^N.$$

**Proof.** This follows directly from Lemma 5.2, from $\langle v/u\rangle < \langle v\rangle/u$ for $0 < u \le 1$ and the relation (1.5) between continuous coefficients and lattice parameters. □

By Lemma 5.6, the term $T_4$ can now be estimated by

$$T_4 \quad \le \quad c_N \cdot a_j^{-p/4} \cdot \sum_{\ell=1}^{a_j^{-1/2}-1} e^{-\rho p\frac{|\sin\theta_{j,\ell}|}{2a_j}} \Big[ \sum_{k\in\mathbf{Z}^2} T_{j,k,\ell}^{Np} \Big].$$

Let $\mathcal{B}_{j,\ell}$ be the rotated anisotropic cartesian grid of curvelet coefficient locations at scale $j$ and orientation $\theta_{j,\ell}$. Note that for $N$ large,

$$a_j^{3/2} \sum_{b \in \mathcal{B}_{j,\ell}} \langle |b_1| \rangle^{-N} \cdot \langle |b_2| \rangle^{-N} \to \int\int \langle |b_1| \rangle^{-N} \cdot \langle |b_2| \rangle^{-N} db_1 db_2, \qquad j \to \infty.$$

Indeed, the function $F(b) = \langle |b_1| \rangle^{-N} \cdot \langle |b_2| \rangle^{-N}$ is smooth and the above display just expresses the fact that Riemann sums of $F$ converge to the integral of $F$. In fact it is quite evident that the convergence is uniform in $\theta$. We conclude that

$$\max_{\ell} \sum_{k \in \mathbf{Z}^2} T_{j,k,\ell}^{Np} \le c \cdot a_j^{-3/2} \cdot a_j^{-Np/2}.$$

We obtain

$$T_4 \le c_{N,p} \cdot a_j^{-p/4} \cdot a_j^{-Np/2} \cdot a_j^{-3/2} \cdot \sum_{\ell=1}^{a_j^{-1/2}-1} e^{-\rho p \frac{|\sin \theta_{j,\ell}|}{2a_j}}.$$

On the interval $0 < \omega \le \pi/2$, $\sin(\omega)/\omega \ge 2/\pi$. We have $|\sin(\pi \ell \sqrt{a_j})| \ge 2\ell\sqrt{a_j}$, $0 \le \ell < a_j^{1/2}/2$. Hence, using $|\sin(\pi/2 + \omega)| = |\sin(\pi/2 - \omega)|$,

$$\sum_{\ell=1}^{a_j^{-1/2}-1} e^{-\rho p \frac{|\sin \theta_{j,\ell}|}{2a_j}} \le 2 \cdot \sum_{\ell=1}^{a_j^{-1/2}/2} e^{-\rho p 2 \ell a_j^{-1/2}}.$$

Summing the geometric series $\sum_{\ell=1}^{\infty} z^\ell$ with $z = e^{-\rho p 2 a_j^{-1/2}}$, we finally obtain that for all $N$ with $Np$ sufficiently large:

$$\sum_{\ell=1}^{a_j^{-1/2}-1} \sum_{k \in \mathbf{Z}^2} |\langle w\mathcal{L}_j, \gamma_{j,k,\ell} \rangle|^p \le c_{N,p} \cdot a_j^{-p/4-Np/2-3/2} \cdot e^{-\rho p 2 a_j^{-1/2}} \le c_{N,p} \cdot a_j^N. \tag{5.8}$$

Summarizing our bounds on $T_1$–$T_4$ and using (5.4), we obtain:

**Lemma 5.7** *For $j$, $N = 1, 2, \ldots$, and $p > 0$, the following holds.*

(i) *We have*

$$\sum_{\{\eta \in \Delta_j^{\pm 1} : (b_{j,k,\ell}, \theta_{j,\ell}) \in \mathcal{N}_2^{PS}(a_j)\}} |\langle w\mathcal{L}_j, \gamma_\eta \rangle|^p \le c_{N,p} \cdot a_j^{-(1/2+\varepsilon)-p/4}.$$

(ii) *We have*

$$\sum_{\{\eta \in \Delta_j^{\pm 1} : (b_{j,k,\ell}, \theta_{j,\ell}) \notin \mathcal{N}_2^{PS}(a_j)\}} |\langle w\mathcal{L}_j, \gamma_\eta \rangle|^p \le c_{N,p} \cdot a_j^{(1-p)/4}.$$

**Proof.** Again reducing to scale $j$ and to the discrete setting as in the proof of Lemma 5.3, (i) follows from $T_1$, i.e., from (5.5). (ii) follows from $T_2$–$T_4$, i.e., from (5.6), (5.7), and (5.8). □

Finally, this lemma now implies that

$$\|\alpha_j\|_p^p \le c_{N,p} \cdot a_j^{-(1/2+\varepsilon)-p/4},$$

which is what was claimed in Lemma 5.3. □

### 5.2 Proof of Lemma 5.4

Using the special properties of our subband filters, WLOG we can assume that $\eta \in \Delta_j$, and can conclude that

$$\|\alpha_j 1_{(\tilde{\mathcal{S}}_j)^c}\|_1 \le c \cdot \sum_{\eta \not\sim \mathcal{N}_2^{PS}(a_j)} |\langle w\mathcal{L}_j, \gamma_\eta \rangle|^p.$$

Applying Lemma 5.7(ii), we obtain

$$\|\alpha_j 1_{(\tilde{\mathcal{S}}_j)^c}\|_1 \le c_{N,1} \cdot a_j^0 = O(1), \qquad j \to \infty.$$

□

## 6 Sparse Expansion of a Curvilinear Singularity

Continuing our 'infrastructure development', we now study properties of curvelet coefficients of a curved singularity. The strategy is to smoothly partition the curve into pieces and then straighten each piece, enabling us to apply results from the previous section.

### 6.1 Tubular Neighborhood

First, we develop a quantitative 'tubular neighborhood theorem'. By regularity, we note that the radius of curvature of $\tau$ is bounded below, by $r > 0$ say. We can find $\rho$ small compared to $r$ and an integer $m$ so that

$$m \cdot \rho = \text{length}(\tau)$$

and so that the integrated curvature of $\tau$ on each interval $[(i-1)\rho, (i+1)\rho]$ is controlled:

$$\int_{(i-1)\rho}^{(i+1)\rho} |\tau''(t)| dt \le \varepsilon. \tag{6.1}$$

Consider the following local coordinate system in the vicinity of $\tau$. Let $t_i = i\rho$, for $i = 0, \ldots, m$, and $T^i = [t_{i-1}, t_{i+1}]$ for $i \in 1, \ldots, m-1$. If $\tau$ is a closed curve, let $T^0 = [t_{m-1}, t_1]$ and $T^m = T^0$ (as $\tau(t_0) \equiv \tau(t_m)$). Let $n_i$ be some choice of unit normal vector to $\tau(t_i)$. For $y \in \mathbf{R}^2$ a point near $\tau(t)$, consider the closest point in $image(\tau)$; this has arclength parameter

$$x_2(y) = \text{argmin}\{|\tau(t) - y| : 0 \le t \le \text{length}(\tau)\}$$

and signed distance parameter

$$x_1^i(y) = \langle n_i, y - \tau(x_2(y)) \rangle \cdot \min\{|\tau(t) - y| : 0 \le t \le \text{length}(\tau)\}.$$

Define the correspondences

$$\phi^i(y) = (x_1^i(y), x_2(y) - t_i), \qquad i = 1, \ldots, m-1,$$

with similar definitions, slightly amended for the case $i \in \{0, m\}$ if $\tau$ is a closed curve. Recall the curvature bound $\varepsilon$ in (6.1).

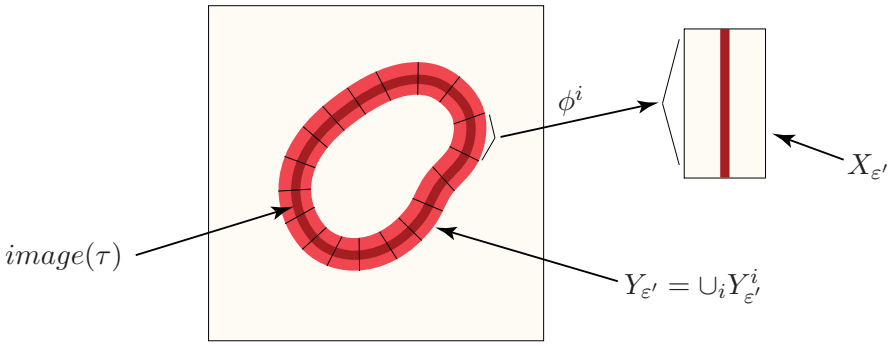


Figure 8: The tubular neighborhood $Y_{\varepsilon'} = \cup_i Y^i_{\varepsilon'}$ of $image(\tau)$ and the mapping $\phi^i : Y^i_{\varepsilon'} \mapsto X_{\varepsilon'}$.

**Lemma 6.1 (Tubular Neighborhood Theorem)** *For sufficiently small $\varepsilon > 0$, there is some $\varepsilon' > 0$ so that, for $X_{\varepsilon'} = [-\varepsilon', \varepsilon'] \times [-\rho, \rho]$, we have:*

- *the correspondence $\phi^i$ is one-one on the set $Y^i_{\varepsilon'} \equiv (\phi^i)^{-1}[X_{\varepsilon'}]$,*
- *the mapping $\phi^i : Y^i_{\varepsilon'} \mapsto X_{\varepsilon'}$ is a diffeomorphism, and*
- *the mapping $\phi^i$ extends to a diffeomorphism from $\mathbf{R}^2$ to $\mathbf{R}^2$ which reduces to the identity outside a compact set.*

In what follows, $\phi^i$ always denotes the extended diffeomorphism from $\mathbf{R}^2$ to $\mathbf{R}^2$.

The set $Y_{\varepsilon'} = \cup_i Y^i_{\varepsilon'}$ is a tubular neighborhood of $image(\tau)$ on which we have nice local coordinate systems, see Figure 8. This will allow us to locally *bend* the curve $\tau$ into something straight.

## 6.2 Cutting into pieces

Choose a $C^\infty$ function $w_2 : \mathbf{R} \mapsto [0, 1]$ (cf. Section 5) supported in $[-1, 1]$ so that

$$w_2(t/\rho) + w_2((t-1)/\rho) = 1, \qquad -1/2 \le t \le 0,$$

and

$$w_2(t/\rho) + w_2((t+1)/\rho) = 1, \qquad 0 \le t \le 1/2.$$

In addition, we require $w_2$ to satisfy

$$|\hat{w}_2(\omega)| \le c \cdot e^{-|\omega|}, \qquad \omega \in \mathbf{R}. \tag{6.2}$$

Define now a smooth partition of unity of $[0, 1]$ using $w_2$:

$$w_{2,i}(t/\rho) = w_2((t - t_i)/\rho), \qquad 1 \le i \le m - 1,$$

with a modification for $i \in \{0, m\}$ that depends on whether $\tau$ is closed or not. Then

$$\sum_i w_{2,i}(t/\rho) = 1 \qquad \forall t \in [0, 1]. \tag{6.3}$$

This will allow us to chop the curve $\tau$ into something that can be bent.

Now define the distributions

$$\mathcal{C}^i = \int_{t_{i-1}}^{t_{i+1}} w_{2,i}(t/\rho)\delta_{\tau(t)}dt;$$

the partition of unity property (6.3) gives $\sum_i \mathcal{C}^i = \mathcal{C}$.

We note that $\{\tau(t) : t \in T^i\} \subset Y^i_{\varepsilon'}$, and hence $\phi^i$ diffeomorphically straightens the piece of curve $\{\tau(t) : t \in T^i\}$ into the line segment $\{0\} \times [-\rho, \rho]$.

## 6.3 Bending one piece

Now consider a diffeomorphism $\phi : \mathbf{R}^2 \mapsto \mathbf{R}^2$; it acts on the distribution $f$ by change of variables

$$\phi^\star f = f \circ \phi.$$

This action induces a linear transformation on the space of curvelet coefficients. With $\alpha(f)$ the curvelet coefficients of $f$ and $\beta(f)$ the curvelet coefficients of $\phi^\star f$, we obtain a linear operator

$$M_\phi(\alpha(f)) = \beta(f).$$

It is by now well-known that diffeomorphisms preserve sparsity of frame coefficients when the frame is based on parabolic scaling (as with curvelets and shearlets). For example, the following can be derived from Hart Smith's work [38] by a simple atomic decomposition.

**Lemma 6.2** *[8, Theorem 6.1, page 219] For $p > 0$, define the operator quasi-norm*

$$\|M_\phi\|_{Op,p} = \max\left\{ \sup_\eta \|(\langle \gamma_\eta, \phi^\star \gamma_{\eta'} \rangle)_\eta\|_p, \sup_{\eta'} \|(\langle \gamma_\eta, \phi^\star \gamma_{\eta'} \rangle)_{\eta'}\|_p \right\}$$

*Let $\phi$ denote a diffeomorphism that reduces to the identity outside of a compact set. Then for $0 < p \leq 1$,*

$$\|M_\phi\|_{Op,p} < c_p < \infty.$$

Far more detailed and precise results on the invariance of curvelet coefficients under $C^2$ changes of variables, with optimal regularity conditions, were developed by Candès and Demanet [5] – as we will see in the next section.

The $p$-triangle inequality $|a + b|^p \leq |a|^p + |b|^p$ for $p \in (0, 1]$ implies the following:

**Lemma 6.3** *For $p \in (0, 1]$, a vector $\alpha = (\alpha_\eta)_\eta$, and a linear operator $M$,*

$$\|M\alpha\|_p \leq \|M\|_{Op,p}\|\alpha\|_p.$$

### 6.4 Gluing pieces together

Now define

$$\beta_j = (\langle \gamma_\eta, \mathcal{C}_j \rangle)_\eta.$$

From the decomposition $\mathcal{C}_j = \sum_{i=1}^m \mathcal{C}_j^i$ we have

$$\beta_j = \sum_{i=1}^m M_{\phi^i} \alpha_j.$$

This decomposition allows us to relate sparsity of coefficients of the linear singularity to those of the curvilinear singularity:

$$\|\beta_j\|_p \quad \leq \quad m^{1/p} \cdot \left( \max_i \|M_{\phi^i}\|_{Op,p} \right) \cdot \|\alpha_j\|_p.$$

This decomposition will be useful below; however, the above argument, which implies sparsity, will not be enough for our main result, which requires also to know the geometric arrangement of the significant coefficients. The next section develops a much finer estimation approach.

## 7 The cluster $\mathcal{S}_{2,j}$ and its estimates

We finally turn to the definition of the cluster set $\mathcal{S}_{2,j}$ and the decisive estimates

$$\delta_{2,j} = o(\|f_j\|_2), \qquad j \to \infty, \tag{7.1}$$

and

$$\mu_c(\mathcal{S}_{2,j}, \{\gamma_\eta\}; \{\psi_\lambda\}) \to 0, \qquad j \to \infty. \tag{7.2}$$

As explained in Section 2.2, combining these results with the results of Section 4 will complete the proof of Theorem 1.1.

We define the cluster of curvelet coefficients indirectly. We first define $\tilde{\mathcal{S}}_j$, the cluster of significant coefficients of our 'straight' model singularity $w\mathcal{L}_j$; then by cutting, bending, and filtering, we induce a cluster for the curvilinear singularity $\mathcal{C}_j$. Set

$$\tilde{\mathcal{S}}_j = \{(j,k,\ell) \in \Delta_j^{\pm 1} : (b_{j,k,\ell}, \theta_{j,\ell}) \in \mathcal{N}_2^{PS}(a_j)\}. \tag{7.3}$$

Lemma 5.4 shows that this set contains the significant coefficients of $w\mathcal{L}_j$.

Let $M_{F_j} = (\langle \gamma_\eta, F_j \star \gamma_{\eta'} \rangle)_{\eta,\eta'}$ be the filtering matrix associated with the filter $F_j$, and recall the definition of the mapping matrix $M_{(\phi^i)^{-1}}$ from Subsection 6.3. Our analysis will require us to consider their product, hence for the sake of brevity we define $M_j^i$ to be

$$M_j^i = M_{F_j} \cdot M_{(\phi^i)^{-1}}$$

and the entries of this matrix by $M_j^i(\eta, \eta')$. Further, we let $t_{\eta',n}$ denote the amplitude of the $n$'th largest element of the $\eta'$'th column. Also let $n_j = 2^{j\varepsilon}$, where $\varepsilon$ was fixed at the beginning of Section 4. We can think of $\varepsilon$ being arbitrarily small, however for our analysis the condition $\varepsilon < 1/28$ will be sufficient.

Morally, what we would like to do is study a cluster of curvelet coefficients built from the cluster pieces

$$\mathcal{S}_j^i = \{\eta : \eta' \in \tilde{\mathcal{S}}_j \text{ and } |M_j^i(\eta, \eta')| > t_{\eta', n_j}\}.$$

In words, $\mathcal{S}_j^i$ consists of the 'top-$n_j$' curvelet coefficients affected by some significant coefficient in $\tilde{\mathcal{S}}_j$. The overall cluster set would then be made by combing the pieces:

$$\mathcal{S}_{2,j} = \bigcup_i \mathcal{S}_j^i.$$

While this morally explains what we do in this section, it turns out that the exact behavior of $\mathcal{S}_j^i$ and $\mathcal{S}_{2,j}$ defined in this natural manner would be rather delicate. In fact, this section uses a more robust definition of cluster set that is similar in spirit; see (7.6)-(7.7) below. This definition depends on some more sophisticated ideas, which we now develop.

## 7.1 Decay Estimates for the Curvelet Representation of FIO's

We first recall some results from [5] on sparsity of curvelet representations of Fourier Integral Operators (FIO's) and decay estimates of such a representation, which will later on be applied to the matrix $M_j^i$.

In order to state decay estimates of the curvelet representation of FIO's, we first require a notion of distance between two curvelet indices. A suitable distance has first been introduced by Hart Smith in [38]. Our analysis will employ results obtained by Candès and Demanet in their work on the curvelet representation of wave propagators [5], in which they use the following variation of Hart Smith's distance:

$$d_{HS}(\eta, \eta') = |\theta_{j,\ell} - \theta_{j',\ell'}|^2 + |b_k - b_{k'}|^2 + |\langle e_\eta, b_k - b_{k'}\rangle|,$$

where

$$b_k = R_{\theta_{j,\ell}} D_{a_j} k \quad \text{and} \quad e_\eta = (\cos(\theta_{j,\ell}), \sin(\theta_{j,\ell})),$$

and the difference $|\theta_{j,\ell} - \theta_{j',\ell'}|$ is understood to refer to geodesic distance in $\mathbf{P}^1$. In [5], this distance was then extended to derive a distance adapted to discrete curvelet indices, which means, in particular, including the scaling component. For a pair of curvelet indices $\eta = (j, k, \ell)$ and $\eta' = (j', k', \ell)$, this so-called *dyadic-parabolic pseudo-distance* is defined by

$$\omega(\eta, \eta') = 2^{|j-j'|} \left(1 + \min\{2^j, 2^{j'}\} d_{HS}(\eta, \eta')\right).$$

We will require the following property of this pseudo-distance:

**Lemma 7.1** *[5, Prop. 2.2 (3.)] For sufficiently large $N > 0$, there is a constant $c_N > 0$ such that*

$$\sum_{\eta''} \omega(\eta, \eta'')^{-N} \cdot \omega(\eta'', \eta')^{-N} \le c_N \cdot \omega(\eta, \eta')^{-(N-1)}, \qquad \forall \eta, \eta'.$$

Another property which will come in handy is the following estimate:

**Lemma 7.2** *[5, Proof of Thm. 1.1] There exist some $N > 0$ and constant $c_N > 0$ obeying*

$$\sum_{\eta} \omega(\eta, \eta')^{-N} \leq c_N, \qquad \forall \eta, \eta'.$$

Before we can state the next result we have to briefly recall some of the key notions in microlocal analysis. Let $S^\star(\mathbf{R}^2)$ denote the cosphere bundle of $\mathbf{R}^2$ – roughly speaking $\{(b_0, \theta_0) : b_0 \in \mathbf{R}^2, \theta_0 \in \mathbf{P}^1\}$ –, and let $\phi$ be a diffeomorphism of $\mathbf{R}^2$. Then the associated so-called *canonical transformation* $\chi$ maps some element $(b_0, \theta_0)$ of phase space into $\chi(b_0, \theta_0) = (\phi(b_0), \phi^\star \theta_0)$, where $\phi^\star \theta_0$ is the codirection into which the codirection $\theta_0$ based infinitesimally at $b_0$ is mapped under $\phi$. Phrasing it differently, we can say that each diffeomorphism of the base space $\mathbf{R}^2$ induces a diffeomorphism of phase space. Such a canonical transformation induces a mapping of curvelet indices which – abusing notation – we again denote by $\chi$. Since we will consider discrete curvelet coefficients $\eta$, we have to be careful how to define this extension. In fact, we will define the image of $\eta$ to be the closest point using the pseudo-distance $\omega$ to the image of $\eta$ under the canonical transformation. As already remarked in [5], choosing a different neighbor only affects the constants in the key inequalities.

The basic insights about parabolic scaling and FIO's are already present in [38], implying sparsity of FIO's of order 0, as explained in [8]. But utilizing the dyadic-parabolic pseudo-distance, Candès and Demanet derived phase space decay estimates for the curvelet representation of FIO's of each order $m$, which imply sparsity, but also inform about geometry.

**Theorem 7.1** *[5, Thm. 5.1] Let $T$ be a Fourier Integral Operator of order $m$ acting on functions of $\mathbf{R}^2$. Then, for each $N > 0$, there exists some positive constant $c_N$ such that*

$$|\langle \gamma_\eta, T\gamma_{\eta'} \rangle| \leq c_N \cdot 2^{mj'} \omega(\eta, \chi(\eta'))^{-N}. \tag{7.4}$$

*Moreover, for each $0 < p < \infty$, $(\langle \gamma_\eta, T\gamma_{\eta'} \rangle)$ is bounded from $\ell_p$ to $\ell_p$.*

In the sequel we will use the first part of the result for $m = 0$. Let us now turn to the decay estimate of the cluster approximate error $\delta_{2,j}$.

## 7.2 $\mathcal{S}_{2,j}$ offers low approximation error

In this section we give two decisive lemmas which drive our analysis, and define $\mathcal{S}_{2,j}$. From now on $\chi^i$ denotes the extension to curvelet indices of the canonical transformation associated with $(\phi^i)^{-1}$.

**Lemma 7.3** *For any $N > 0$, there exists a positive constant $c_N$ such that*

$$|M_j^i(\eta, \eta')| \leq c_N \cdot \omega(\eta, \chi^i(\eta'))^{-N}, \qquad \forall \eta, \eta'. \tag{7.5}$$

**Proof.** Given $N > 0$, by Theorem 7.1, there exists some positive constant $c_N$ such that (7.4) holds, which implies both

$$|\langle \gamma_\eta, F_j \star \gamma_{\eta'} \rangle| \leq c_N \cdot \omega(\eta, \eta')^{-(N+1)} \quad \text{and} \quad |\langle \gamma_\eta, \gamma_{\eta'} \circ (\phi^i)^{-1} \rangle| \leq c_N \cdot \omega(\eta, \chi^i(\eta'))^{-(N+1)}.$$

Now applying Lemma 7.1 proves the claim. □

**Lemma 7.4** *There is a constant $c_1 > 0$ such that for each vector $\alpha = (\alpha_\eta)_\eta$ ,*

$$\|M_j^i \alpha\|_1 \le c_1 \cdot \|\alpha\|_1.$$

**Proof.** This already follows from Lemma 6.3. However, it is instructive to reprove it using (7.5) of Lemma 7.3:

$$\|M_j^i \alpha\|_1 \le \sup_{\eta'} \sum_{\eta} |M_j^i(\eta, \eta')| \cdot \|\alpha\|_1 \le c_N \cdot \sup_{\eta'} \sum_{\eta} \omega(\eta, \chi^i(\eta'))^{-N} \cdot \|\alpha\|_1.$$

Realizing that the $\sup_{\eta'}$ allows us to omit $\chi^i$ and applying Lemma 7.2,

$$\sup_{\eta'} \sum_{\eta} \omega(\eta, \chi^i(\eta'))^{-N} = \sup_{\eta'} \sum_{\eta} \omega(\eta, \eta')^{-N} \le c_N. \qquad \square$$

These two lemmas say that, in place of studying $M_j^i$ and its detailed properties, we can simply study its majorant $c_N \cdot \omega(\eta, \chi^i(\eta'))^{-N}$. So fix $N$ large and let $\tilde{M}_j^i$ denote the 'model'

$$\tilde{M}_j^i(\eta, \eta') = c_N \cdot \omega(\eta, \chi^i(\eta'))^{-N}. \tag{7.6}$$

We define our cluster set $\mathcal{S}_{2,j}$ in terms of the model rather than in terms of $M_j^i$, via

$$\mathcal{S}_{2,j} = \bigcup_i \mathcal{S}_j^i, \tag{7.7}$$

where

$$\mathcal{S}_j^i = \{\eta : \eta' \in \tilde{\mathcal{S}}_j \text{ and } |\tilde{M}_j^i(\eta, \eta')| > t_{\eta', n_j}\}.$$

In this definition, $\mathcal{S}_j^i$ is not truly the set of significant coefficients, but rather a set of sites where significant coefficients could *potentially* occur, given the geometry of the problem; so it is a bit larger. We still speak of $\mathcal{S}_{2,j}$ as if it were exactly the set of significant coefficients.

The set $\tilde{\mathcal{S}}_j$ is explicitly defined by a tube in phase space; the tube becomes narrower at finer scales and 'converges' to $WF(w\mathcal{L})$. The set of potentially significant coefficients $\mathcal{S}_j^i$ is a much thicker set and gets progressively thicker relative to $\tilde{\mathcal{S}}_j$ with increasing $j$, however, geometrically the corresponding 'tube' is still becoming very narrow as $j$ increases. This device already appeared in the Heuristics section; it allows to conveniently bound all the insignificant interactions $M_j^i(\eta, \eta')$; in particular, see the estimate of $T_2$ in the proof of Lemma 7.5.

We can now prove the estimate (7.1) for the cluster approximate error $\delta_{2,j}$.

**Lemma 7.5**

$$\delta_{2,j} = o(\|f_j\|_2), \qquad j \to \infty.$$

**Proof.** As in Section 6.4, let $\beta_j = (\langle \gamma_\eta, \mathcal{C}_j \rangle)_\eta$ as well as $\beta_j^i = (\langle \gamma_\eta, \mathcal{C}_j^i \rangle)_\eta$, $i \in \{0, \ldots, m-1\}$. The decomposition $\mathcal{C}_j = \sum_i \mathcal{C}_j^i$ implies $\beta_j = \sum_i \beta_j^i$. Now

$$\begin{aligned} \sum_{\eta \in \Delta \setminus \mathcal{S}_{2,j}} |\beta_j(\eta)| &\le \sum_i \sum_{\eta \in \Delta_j^{\pm 1} \setminus \mathcal{S}_j^i} |\beta_j^i(\eta)| \\ &\le m \cdot \max_i \sum_{\eta \in \Delta_j^{\pm 1} \setminus \mathcal{S}_j^i} |\beta_j^i(\eta)| \le m \cdot \max_i \|\beta_j^i \cdot 1_{\Delta_j^{\pm 1} \setminus \mathcal{S}_j^i}\|_1. \end{aligned}$$

We now decompose $\|\beta_j^i \cdot 1_{\Delta_j^{\pm 1} \setminus \mathcal{S}_j^i}\|_1$ into three components and estimate each separately. Let $u_j := \lceil \varepsilon \cdot j \rceil$ which will be the 'radius' of the scale-neighborhood about scale $j$ that we distinguish from the remaining (unimportant) scales. In what follows, remember the definition of $\tilde{\mathcal{S}}_j$ in (7.3); and let $\tilde{\mathcal{S}}_j^{\pm u_j} = \bigcup_{j'} \tilde{\mathcal{S}}_{j'}$ . Then

$$\begin{aligned}
&\|\beta_j^i \cdot 1_{\Delta_j^{\pm 1} \setminus \mathcal{S}_j^i}\|_1 \\
&\quad \leq \quad \|\tilde{M}_j^i(\alpha \cdot 1_{\Delta_j^{\pm u_j} \setminus \tilde{\mathcal{S}}_j^{\pm u_j}}) \cdot 1_{\Delta_j^{\pm 1} \setminus \mathcal{S}_j^i}\|_1 + \|\tilde{M}_j^i(\alpha \cdot 1_{\tilde{\mathcal{S}}_j^{\pm u_j}}) \cdot 1_{\Delta_j^{\pm 1} \setminus \mathcal{S}_j^i}\|_1 \\
&\qquad + \|\tilde{M}_j^i(\alpha \cdot 1_{\Delta \setminus \Delta_j^{\pm u_j}}) \cdot 1_{\Delta_j^{\pm 1} \setminus \mathcal{S}_j^i}\|_1 \\
&\quad = \quad T_1 + T_2 + T_3. \qquad (7.8)
\end{aligned}$$

First, let's estimate $T_1$. From Lemma 7.4, we obtain

$$T_1 \leq \sum_{j'=j-u_j}^{j+u_j} \|\tilde{M}_j^i(\alpha \cdot 1_{\Delta_{j'} \setminus \tilde{\mathcal{S}}_{j'}})\|_1 \leq c_1 \cdot \sum_{j'=j-u_j}^{j+u_j} \|\alpha \cdot 1_{\Delta_{j'} \setminus \tilde{\mathcal{S}}_{j'}}\|_1.$$

By Lemma 5.4,

$$\|\alpha \cdot 1_{\Delta_{j'} \setminus \tilde{\mathcal{S}}_{j'}}\|_1 = O(1), \qquad j' \to \infty;$$

hence,

$$T_1 \leq c_1 \cdot (2u_j + 1) = O(2^{j\varepsilon}), \qquad j \to \infty. \qquad (7.9)$$

Next, turn to $T_2$. Observe that

$$T_2 \leq \sum_{j'=j-u_j}^{j+u_j} \Big[ \sup_{\eta' \in \tilde{\mathcal{S}}_{j'}} \sum_{\eta \in \Delta_j^{\pm 1} \setminus \mathcal{S}_j^i} |\tilde{M}_j^i(\eta, \eta')| \Big] \cdot \|\alpha \cdot 1_{\tilde{\mathcal{S}}_{j'}}\|_1. \qquad (7.10)$$

We now need the following standard lemma about $n$-term approximations.

**Lemma 7.6** *Let $x = (x_i)_i$ denote a sequence of numbers and let $|x|_{(n)}$ denote the nth-largest element in the decreasing rearrangement. For $0 < p < 1$ we have the inequality:*

$$\sum_i |x_i| \cdot 1_{\{|x_i| \leq |x|_{(n)}\}} \leq c_p \cdot \|x\|_p \cdot n^{-(1-p)/p}, \quad n = 1, 2, \ldots.$$

Recalling that $\mathcal{S}_j^i$ consists of elements $\eta$ such that $|\tilde{M}_j^i(\eta, \eta')| > t_{\eta', n_j}$, we conclude

$$\sup_{\eta' \in \tilde{\mathcal{S}}_{j'}} \sum_{\eta \in \Delta_j^{\pm 1} \setminus \mathcal{S}_j^i} |\tilde{M}_j^i(\eta, \eta')| \leq c_p \cdot \left( \sup_{\eta'} \sum_{\eta} |\tilde{M}_j^i(\eta, \eta')|^p \right)^{1/p} \cdot n_j^{-(1-p)/p} \leq c_1 \cdot c_p \cdot n_j^{-(1-p)/p}.$$

Also, from Lemma 5.3, we obtain $\|\alpha \cdot 1_{\tilde{\mathcal{S}}_{j'}}\|_1 \leq \|\alpha \cdot 1_{\Delta_{j'}}\|_1 \leq \tilde{c}_1 2^{j'(3/4+\epsilon)}$. Choose $p$ so that $p < 4\varepsilon/(3 + 7\varepsilon)$; returning to (7.10),

$$T_2 \leq c \cdot n_j^{-(1-p)/p} \sum_{j'=j-u_j}^{j+u_j} 2^{j'(3/4+\epsilon)} \leq c \cdot 2^{-(1-p)\varepsilon j/p} \cdot (2u_j + 1) \cdot 2^{3(j+u_j)/4 + j\epsilon} = O(2^{j\varepsilon}), \; j \to \infty, \qquad (7.11)$$

At last, we consider $T_3$. Notice that

$$T_3 \le \sum_{|j-j'|\ge u_j} \Big[ \sup_{\eta'\in\Delta_{j'}} \sum_{\eta\in\Delta_j^{\pm 1}\setminus\mathcal{S}_j^i} |\tilde{M}_j^i(\eta,\eta')| \Big] \cdot \|\alpha\cdot 1_{\Delta_{j'}}\|_1. \tag{7.12}$$

Using the definition of $\tilde{M}_j$, we proceed as in the proof of Lemma 7.4, and employ the definition of the pseudo-distance $\omega$; we'll obtain

$$\begin{aligned} \sup_{\eta'\in\Delta_{j'}} \sum_{\eta\in\Delta_j^{\pm 1}\setminus\mathcal{S}_j^i} |\tilde{M}_j^i(\eta,\eta')| &\le \tilde{c}_N \cdot \sup_{\eta'\in\Delta_{j'}} \sum_{\eta\in\Delta_j^{\pm 1}\setminus\mathcal{S}_j^i} \omega(\eta,\eta')^{-N} \\ &\le c\cdot 2^{-N|j-j'|}. \end{aligned}$$

By Lemma 5.3, the second term in (7.12) can be estimated by $\|\alpha\cdot 1_{\Delta_{j'}}\|_1 \le c\cdot 2^{j'(3/4+\epsilon)}$. We conclude that for $N$ sufficiently large,

$$T_3 \le c_N \sum_{|j-j'|\ge u_j} 2^{j'(3/4+\epsilon)-N|j-j'|} = O(2^{j\varepsilon}),\ j\to\infty. \tag{7.13}$$

Combining (7.9), (7.11), and (7.13) with (7.8) yields

$$\delta_{2,j} = \sum_{\eta\in(\mathcal{S}_j^i)^c} |\beta_j^i(\eta)| = O(2^{j\varepsilon}) = o(\|f_j\|_2), \qquad j\to\infty. \quad \square$$

### 7.3 $\mathcal{S}_{2,j}$ offers low cluster coherence

This section proves (7.2), the asymptotically negligible cluster coherence of $\mathcal{S}_{2,j}$.

**Lemma 7.7** *We have*

$$\mu_c(\mathcal{S}_{2,j}, \{\gamma_\eta\}; \{\psi_\lambda\}) \to 0, \qquad j\to\infty.$$

Before giving the proof, we state two useful lemmas. Both use the variables $n_j$ introduced earlier. The first lemma implies that a given significant curvelet coefficient in the analysis of $w\mathcal{L}_j$ pushes forward to produce significant coefficients at roughly the same scale, and near a certain fixed orientation and location. Thus the pushforward acts roughly like a rigid motion. That first lemma is proved in Section 9.5.1.

**Definition 7.1** *Let the canonical transformation $\chi$ be given. For a specific curvelet index $\eta$ the* forward set *of radius $n$ is:*

$$\textsc{fwd}(\eta;\chi,n) = \{\eta' : \omega(\eta',\chi(\eta)) \le n\}.$$

In words, FWD is the set of curvelet indices close to the pushforward of $\eta$ by $\chi$. Note that the forward set covers the set of significant interactions with $\eta$:

$$\{\eta' : |\tilde{M}_j^i(\eta',\eta)| > t_{\eta,n_j}\} \subset \textsc{fwd}(\eta;\chi^i,n_j).$$

Consequently

$$\mathcal{S}_{2,j} \subset \bigcup_{\eta\in\tilde{\mathcal{S}}_j} \bigcup_{i\in\{0,\dots,m\}} \textsc{fwd}(\eta;\chi^i,n_j).$$

**Lemma 7.8** *Let $\eta = (j,k,\ell)$ be a curvelet index, with its image under the canonical transformation $\chi^i$ denoted by $\chi^i(\eta) = (\tilde{j},\tilde{k},\tilde{\ell})$ for a fixed $i \in \{0,\dots,m\}$. There exists some positive constant $c > 0$ such that, for $j$ sufficiently large,*

$$\begin{aligned} \text{FWD}(\eta;\chi^i,n_j) \quad \subseteq \quad & \{(j',k',\ell') : |j'-\tilde{j}| \le c\cdot \log n_j,\ |2^{(j'-\tilde{j})/2}\tilde{\ell} - \ell'|^2 \le c\cdot n_j, \\ & 2^{j'}\cdot \max\{|b_{\tilde{k}} - b_{k'}|^2, |b_{\tilde{k}} - b_{k'}|\} \le c\cdot n_j\} \end{aligned}$$

We conclude that for sufficiently large $j_0$, there exists $c > 0$ so that

$$\#\text{FWD}(\eta;\chi^i,n_j) \le c\cdot n_j^4, \qquad j > j_0, \quad \eta \in \tilde{\mathcal{S}}_j.$$

Let $k(\eta)$ denote the $k$-component of $\eta = (j,k,\ell)$. Define

$$d_{min}(\eta,i,n_j) = \min\{|k(\eta')| : \eta' \in \text{FWD}(\eta;\chi^i,n_j)\}.$$

We also need the fact that points in the forward set of $\eta$ have spatial components almost as far from the origin as the spatial component of $\eta$ itself.

**Lemma 7.9** *There are $j_0, c_{1,0}, c_{2,0}$ so that we have*

$$d_{min}(\eta;i,n_j) \ge c_1(j)|k(\eta)| - c_2(j), \qquad j > j_0,$$

*where*

$$c_1(j) = c_{1,0}/n_j \quad \text{and} \quad c_2(j) = c_{2,0}\cdot n_j.$$

The proof is given in the appendix, as is the proof of

**Lemma 7.10** *For $N > 2$, there are constants $c_3$,$c_4$,$c_5$ such that*

$$\sum_{k\in\mathbf{Z}^2} \langle (a|k| - b)_+\rangle^{-N} \le (b/a)^2\cdot(c_3 + c_4 b^{-N}) + c_5.$$

With the last three lemmas we can now prove the main result of this section.

**Proof of Lemma 7.7.** The definition of $\mathcal{S}_{2,j}$ implies

$$\begin{aligned} \mu_c(\mathcal{S}_{2,j},\{\gamma_\eta\};\{\psi_\lambda\}) \quad &= \quad \max_{j',k'} \sum_{\eta\in\mathcal{S}_{2,j}} |\langle \gamma_\eta, \psi_{j',k'}\rangle| = \max_{j',k'} \sum_{i=0}^{m-1} \sum_{\eta\in\mathcal{S}_j^i} |\langle \gamma_\eta, \psi_{j',k'}\rangle| \\ &\le \quad m\cdot \max_i \max_{j',k'} \sum_{\eta\in\mathcal{S}_j^i} |\langle \gamma_\eta, \psi_{j',k'}\rangle| \end{aligned}$$

WLOG assume that $j' = j$ and $k' = 0$, reducing the task to proving that $\sum_{\eta\in\mathcal{S}_j^i} |\langle \gamma_\eta, \psi_{j,0}\rangle| \to 0$ as $j\to\infty$ for all $i$. We have the estimate

$$\mu_c(\mathcal{S}_{2,j},\{\gamma_\eta\};\{\psi_\lambda\}) \le m\cdot \max_i \sum_{\eta\in\mathcal{S}_j^i} |\langle \gamma_\eta, \psi_{j,0}\rangle|. \tag{7.14}$$

We can now continue (7.14) by applying Lemma 3.3 and the three lemmas immediately above:

$$\begin{aligned}
\sum_{\eta\in\mathcal{S}_j^i} |\langle \gamma_\eta, \psi_{j,0}\rangle| &\le \sum_{\eta\in\tilde{\mathcal{S}}_j} \sum_{\eta'\in \mathrm{FWD}(\eta)} |\langle \gamma_{\eta'}, \psi_{j,0}\rangle| \\
&\le \sum_{\eta\in\tilde{\mathcal{S}}_j} \sum_{\eta'\in \mathrm{FWD}(\eta)} c_N \cdot 2^{-j/4} \cdot \langle |k(\eta')|\rangle^{-N} \\
&\le c \cdot 2^{-j/4} \cdot \sum_{\eta\in\tilde{\mathcal{S}}_j} \#\mathrm{FWD}(\eta) \cdot \langle d_{min}(\eta)\rangle^{-N} \\
&\le c \cdot 2^{-j/4} \cdot n_j^4 \cdot \sum_{\eta\in\tilde{\mathcal{S}}_j} \langle (c_1(j)|k(\eta)| - c_2(j))_+\rangle^{-N} \\
&\le c \cdot 2^{-j/4} \cdot n_j^4 \cdot \left( \frac{c_2(j)^2}{c_1(j)^2} \cdot (c_3 + c_4 \cdot c_1(j)^2 (c_2(j))^{-N}) + c_5 \right) \\
&\le c \cdot 2^{-j/4} \cdot n_j^8, \qquad j > j_0.
\end{aligned}$$

We have assumed that $\varepsilon < 1/32$, so $2^{-j/4} n_j^8 = 2^{-j/4} \cdot 2^{8\varepsilon j} \to 0$ as $j \to \infty$; substituting this into (7.14) proves the lemma. □

# 8 Discussion

## 8.1 Extensions

So far we focused entirely on a very special separation problem using very specific tools of harmonic analysis. Our goal was to show that a certain set of questions and results make sense and provide insight. This is the 'tip of the iceberg': the main results are susceptible of very extensive generalizations and extensions.

- *More General Classes of Objects.* We may vary the problem, taking point and curve singularities whose 'strength' is different than the ones we chose in (1.1)-(1.2); however, always matching the strength of the point singularity to that of the curve singularity. For example, consider a 'cartoon' image model, where $\mathcal{C}$ is a function smooth away from discontinuities, and the components of the continuity set are bounded by a complex of smooth curves. Such cartoons still exhibit curvilinear singularities, but the singularities are of order zero rather than order $-1$. For a separation problem with nontrivial asymptotics, we replace the point singularity $|x - x_i|^{-3/2}$ in (1.1) by $|x - x_i|^{-1/2}$, preserving an energy-matching condition like (1.3), with $r^{-1}$ replacing $r^1$. Recall that, without energy matching, the whole problem is trivial. With such changes the proof of Theorem 1.1 will run very closely in parallel. As a general rule, if $\langle \mathcal{C}, f\rangle = \int (\triangle^\gamma f)(\tau(t))dt$, where $\gamma$ is a fractional power of the Laplacian $\triangle$, then matching point singularities have strength $\alpha = (-3 + 4\gamma)/2$. The case we studied in this paper was $\gamma = 0$ and hence $\alpha = -3/2$.

- *Other Frame Pairs.* Theorem 1.1 holds without change for many other pairs of frames and bases. Consider this pair:

- ◇ *Orthonormal Separable Meyer Wavelets* – an orthonormal basis of perfectly isotropic generating elements.
- ◇ *Shearlets* – a highly directional tight frame with increasingly anisotropic elements at fine scales.

In this pair, the wavelets are actually orthonormal, and both wavelets and shearlets correspond very closely to discrete transforms used in digital image processing. In digital image processing, the notions of 'radial', 'directional', 'rotation' and so on are problematic; both orthonormal wavelets and shearlets avoid such concepts. At the same time this pair offers the same ability to sparsify point and curve singularities as the counterparts pair we introduced above. This allows to provide a complete methodology for the continuous and discrete setting (see, e.g., [28, 31]) as well as for algorithmic realizations (see, e.g., [33, 32]).

While the proof arguments explicitly cover the one frame pair we have taken pains to define so far, those arguments extend immediately to other 'compatible' pairs – where the cross-frame matrices are almost diagonal in a suitable sense. This grants us the freedom to prove results in one system which is convenient, but apply those to another compatible system. The arguments showing that shearlets and curvelets are compatible are supplied in [21]. In this paper we discussed the pair radial wavelets/curvelets. However, all results hold true in a similar way for the pair orthonormal wavelets/shearlets.

- *Noisy Data.* Are the results studied here robust against small modelling errors? In fact they are. Consider an image composed of $\mathcal{P}$ and $\mathcal{C}$ with additive noise $\mathcal{N}$, hence we measure $\tilde{f} = \mathcal{P} + \mathcal{C} + \mathcal{N}$ instead of $f$. We then – as in the noiseless case – filter to obtain subband components $\tilde{f}_j = \mathcal{P}_j + \mathcal{C}_j + \mathcal{N}_j$ and apply (CSep) to $\tilde{f}_j$ to obtain a pair $(\tilde{W}_j, \tilde{C}_j)$. Provided that the noise component $\mathcal{N}$ has 'sufficiently' small curvelet coefficients in the sense that at each scale $j$ the $\ell_1$ norm of the analysis coefficients satisfies $o(2^{j/2})$ as $j \to \infty$, we again obtain asymptotically perfect separation:

$$\frac{\|\tilde{W}_j - \mathcal{P}_j\|_2 + \|\tilde{C}_j - \mathcal{C}_j\|_2}{\|\mathcal{P}_j\|_2 + \|\mathcal{C}_j\|_2} \to 0, \qquad j \to \infty$$

This can be proved along the lines of the proof of Theorem 1.1. Indeed, consider a composed signal $\tilde{S} = S_1^0 + S_2^0 + n$ with components $S_1^0$ and $S_2^0$ relatively sparse as in Proposition 2.1, and noise term $n$ satisfying $\|\Phi_1^T n\|_1 < \varepsilon$ or $\|\Phi_2^T n\|_1 < \varepsilon$. Let $(\tilde{S}_1^\star, \tilde{S}_2^\star)$ solve (2.1) with $S$ substituted by $\tilde{S}$. Then following the proof of Proposition 2.1 line by line and adapting the arguments accordingly shows:

$$\|\tilde{S}_1^\star - S_1^0\|_2 + \|\tilde{S}_2^\star - S_2^0\|_2 \leq \frac{2\delta + 5\varepsilon}{1 - 2\kappa}. \tag{8.1}$$

Substituting Proposition 2.1 by estimate (8.1) in the proof of Theorem 1.1 implies the result on geometric separation of noisy data stated in Subsection 8.1.

We conclude that our analysis is indeed stable.

- *Rate of Convergence.* One might wonder about the rate of separation. The lemmas proven in Sections 4-7 imply the following upper bound on the rate of convergence for $\ell_1$ minimization:

$$\frac{\|W_j - \mathcal{P}_j\|_2 + \|C_j - \mathcal{C}_j\|_2}{\|\mathcal{P}_j\|_2 + \|\mathcal{C}_j\|_2} = O(2^{-j(1/2-\epsilon)}), \qquad j \to \infty.$$

  Such information might be the key to getting even stronger separation conclusions.

- *Other Algorithms and Other Notions of Separation.* In the companion paper [20] we show that one pass of alternating hard thresholding, properly tuned, can achieve asymptotic separation. Surprisingly, we can even show clean separation at the level of wavefront sets.

## 8.2 Interpretation as an Uncertainty Principle

Separation results such as the 'birth problem' of $\ell_1$ component separation, a combination of sinusoids and spikes [18], have been interpreted at that time as uncertainty principles. As a reminder to the reader, the classical uncertainty principle states that a signal cannot be highly concentrated in both time and frequency; and a lower bound is placed on the product of the concentration in time and in frequency. The core property which allows the separation of sinusoids and spikes by using a dictionary consisting of the unit basis and the Fourier basis, is the non-existence of a sparse representation of a signal both in time and in frequency.

Considering the present separation problem, these core ideas need to be extended, thereby providing us with yet another interpretation than the one already presented in the previous sections. The two representation 'domains' are now the isotropic system of wavelets and the anisotropic system of curvelets. Hence, we might regard the separation result of Theorem 1.1 as a statement that a 2D Schwartz distribution cannot be sparsely represented via analysis coefficients both in the 'isotropic world' and in the 'anisotropic world'. In particular, if a 2D Schwartz distribution has a sparse representation in wavelets, it is not sparse in curvelets and vice versa. Phrasing it in more general terms, a 2D Schwartz distribution having only isotropic features cannot be sparsely represented using an anisotropic system, and if it has only exhibits anisotropic phenomena, it does not possess a sparse representation in terms of an isotropic system.

Summarizing, comparison with the classical uncertainty principle shows that we here derive an uncertainty principle for the isotropy-anisotropy relation instead of the classical time-frequency relation.

# 9 Proofs

## 9.1 Proofs of Results from Section 2

### 9.1.1 Proof of Proposition 2.1

**Proof.** Since $\Phi_1$ and $\Phi_2$ are tight frames,

$$\begin{aligned}\|S_1^\star - S_1^0\|_2 + \|S_2^\star - S_2^0\|_2 &= \|\Phi_1^T(S_1^\star - S_1^0)\|_2 + \|\Phi_2^T(S_2^\star - S_2^0)\|_2 \\ &\leq \|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_2^\star - S_2^0)\|_1.\end{aligned}$$

Now invoke exact decomposition: $S_1^0 + S_2^0 = S = S_1^\star + S_2^\star$. Rewrite the last display:

$$\|S_1^\star - S_1^0\|_2 + \|S_2^\star - S_2^0\|_2 \leq \|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_1^\star - S_1^0)\|_1.$$

By definition of $\kappa$,

$$\begin{aligned}&\|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_1^\star - S_1^0)\|_1 \\ &\quad = \|1_{\mathcal{S}_1}\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_2}\Phi_2^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_1^c}\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T(S_2^\star - S_2^0)\|_1 \\ &\quad \leq \kappa \cdot \left(\|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_1^\star - S_1^0)\|_1\right) + \|1_{\mathcal{S}_1^c}\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T(S_2^\star - S_2^0)\|_1;\end{aligned}$$

use relative sparsity of the subsignals $S_i^0$, $i = 1, 2$,

$$\begin{aligned}&\|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_1^\star - S_1^0)\|_1 \\ &\quad \leq \frac{1}{1-\kappa}(\|1_{\mathcal{S}_1^c}\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T(S_2^\star - S_2^0)\|_1) \\ &\quad \leq \frac{1}{1-\kappa}(\|1_{\mathcal{S}_1^c}\Phi_1^T S_1^\star\|_1 + \|1_{\mathcal{S}_1^c}\Phi_1^T S_1^0\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T S_2^\star\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T S_2^0\|_1) \\ &\quad \leq \frac{1}{1-\kappa}\left(\|1_{\mathcal{S}_1^c}\Phi_1^T S_1^\star\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T S_2^\star\|_1 + \delta\right). \qquad (9.1)\end{aligned}$$

Apply minimality of $S_1^\star$ and $S_2^\star$,

$$\begin{aligned}\|1_{\mathcal{S}_1^c}\Phi_1^T S_1^\star\|_1 + \|1_{\mathcal{S}_1}\Phi_1^T S_1^\star\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T S_2^\star\|_1 + \|1_{\mathcal{S}_2}\Phi_2^T S_2^\star\|_1 &= \|\Phi_1^T S_1^\star\|_1 + \|\Phi_2^T S_2^\star\|_1 \\ &\leq \|\Phi_1^T S_1^0\|_1 + \|\Phi_2^T S_2^0\|_1.\end{aligned}$$

Again use sparsity of the subsignals $S_i^0$, $i = 1, 2$,

$$\begin{aligned}&\|1_{\mathcal{S}_1^c}\Phi_1^T S_1^\star\|_1 + \|1_{\mathcal{S}_2^c}\Phi_2^T S_2^\star\|_1 \\ &\quad \leq \|\Phi_1^T S_1^0\|_1 + \|\Phi_2^T S_2^0\|_1 - \|1_{\mathcal{S}_1}\Phi_1^T S_1^\star\|_1 - \|1_{\mathcal{S}_2}\Phi_2^T S_2^\star\|_1 \\ &\quad \leq \|\Phi_1^T S_1^0\|_1 + \|\Phi_2^T S_2^0\|_1 + \|1_{\mathcal{S}_1}\Phi_1^T(S_1^\star - S_1^0)\|_1 - \|1_{\mathcal{S}_1}\Phi_1^T S_1^0\|_1 \\ &\qquad + \|1_{\mathcal{S}_2}\Phi_2^T(S_2^\star - S_2^0)\|_1 - \|1_{\mathcal{S}_2}\Phi_2^T S_2^0\|_1 \\ &\quad \leq \|1_{\mathcal{S}_1}\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_2}\Phi_2^T(S_2^\star - S_2^0)\|_1 + \delta.\end{aligned}$$

Using (9.1), this leads to

$$\begin{aligned}&\|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_1^\star - S_1^0)\|_1 \\ &\quad \leq \frac{1}{1-\kappa}\left[\|1_{\mathcal{S}_1}\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|1_{\mathcal{S}_2}\Phi_2^T(S_1^\star - S_1^0)\|_1 + 2\delta\right] \\ &\quad \leq \frac{1}{1-\kappa}\left[\kappa \cdot (\|\Phi_1^T(S_1^\star - S_1^0)\|_1 + \|\Phi_2^T(S_1^\star - S_1^0)\|_1) + 2\delta\right].\end{aligned}$$

Thus, finally we obtain

$$\|S_1^\star - S_1^0\|_2 + \|S_2^\star - S_2^0\|_2 \leq \left(1 - \frac{\kappa}{1-\kappa}\right)^{-1} \cdot \frac{2\delta}{1-\kappa} = \frac{2\delta}{1-2\kappa}.\square$$

### 9.1.2 Proof of Lemma 2.1

**Proof.** For each $f$, we choose coefficient sequences $\alpha_1$ and $\alpha_2$ such that $f = \Phi_1\alpha_1 = \Phi_2\alpha_2$ and $\|\alpha_i\|_1 \leq \|\beta_i\|_1$ for all $\beta_i$ satisfying $f = \Phi_i\beta_i$, $i = 1, 2$. Then, employing the fact that, because $\Phi_1$ and $\Phi_2$ are tight frames, also $f = \Phi_i\Phi_i^T\Phi_i\alpha_i$, $i = 1, 2$, we obtain

$$\begin{aligned}
&\|1_{\mathcal{S}_1}\Phi_1^T f\|_1 + \|1_{\mathcal{S}_2}\Phi_2^T f\|_1 \\
&= \|1_{\mathcal{S}_1}\Phi_1^T\Phi_2\alpha_2\|_1 + \|1_{\mathcal{S}_2}\Phi_2^T\Phi_1\alpha_1\|_1 \\
&\leq \sum_{i\in\mathcal{S}_1}\left(\sum_j |\langle\phi_{1,i},\phi_{2,j}\rangle||\alpha_{2,j}|\right) + \sum_{j\in\mathcal{S}_2}\left(\sum_i |\langle\phi_{1,i},\phi_{2,j}\rangle||\alpha_{1,i}|\right) \\
&= \sum_j\left(\sum_{i\in\mathcal{S}_1}|\langle\phi_{1,i},\phi_{2,j}\rangle|\right)|\alpha_{2,j}| + \sum_i\left(\sum_{j\in\mathcal{S}_2}|\langle\phi_{1,i},\phi_{2,j}\rangle|\right)|\alpha_{1,i}| \\
&\leq \mu_c(\mathcal{S}_1,\Phi_1;\Phi_2)\|\alpha_2\|_1 + \mu_c(\mathcal{S}_2,\Phi_2;\Phi_1)\|\alpha_1\|_1 \\
&\leq \max\{\mu_c(\mathcal{S}_1,\Phi_1;\Phi_2),\mu_c(\mathcal{S}_2,\Phi_2;\Phi_1)\}(\|\alpha_1\|_1 + \|\alpha_2\|_1) \\
&\leq \max\{\mu_c(\mathcal{S}_1,\Phi_1;\Phi_2),\mu_c(\mathcal{S}_2,\Phi_2;\Phi_1)\}(\|\Phi_1^T\Phi_1\alpha_1\|_1 + \|\Phi_2^T\Phi_2\alpha_2\|_1) \\
&= \max\{\mu_c(\mathcal{S}_1,\Phi_1;\Phi_2),\mu_c(\mathcal{S}_2,\Phi_2;\Phi_1)\}(\|\Phi_1^T f\|_1 + \|\Phi_2^T f\|_1).
\end{aligned}$$

$\square$

## 9.2 Proofs of Results from Section 3

### 9.2.1 Proof of Lemma 3.3

Using Parseval, $\langle\gamma_{a,b,\theta},\psi_{a_0,b_0}\rangle = 2\pi\int\hat{\gamma}_{a,b,\theta}(\xi)\hat{\psi}_{a_0,b_0}(\xi)d\xi$, we consider

$$\int\hat{\gamma}_{a,b,\theta}(\xi)\hat{\psi}_{a_0,b_0}(\xi)d\xi \;= \int a_0 W(a_0 r)e^{-ib_0'\xi}\cdot a^{3/4}W(ar)V((\omega-\theta)/\sqrt{a})e^{-ib'\xi}d\xi.$$

Now WLOG we may consider the special case $\theta = 0$, so that $R_\theta = I$. Recall that $W$ is supported on $[1/2, 2]$ by construction. Then

$$W(a_0 r)W(ar) = 0 \qquad \forall r \geq 0, |\log_2(a/a_0)| \geq 3.$$

Hence we need only consider the case where $|\log_2(a/a_0)| < 3$, and in that circumstance we may WLOG take $a = a_0$. We may also assume $b_0 = 0$. Apply the change of variables $\zeta = D_a\xi$ and $d\zeta = a^{3/2}d\xi$,

$$\int\hat{\gamma}_{a,b,\theta}(\xi)\hat{\psi}_{a,b_0}(\xi)d\xi = a^{1/4}\cdot\int W^2(\|\zeta_a\|)V(\omega(\zeta_a)/\sqrt{a})e^{-i(D_{1/a}b)'\zeta}d\zeta,$$

where $\zeta_a = (\zeta_1, \sqrt{a}\zeta_2)$ and $\omega(\zeta_a)$ denotes the angular component of the polar coordinates of $\zeta_a$. Applying integration by parts, for any $k = 1, 2, ...,$

$$\begin{aligned} |\langle \gamma_{a,b,\theta}, \psi_{a_0,b_0} \rangle| &= 2\pi \cdot a^{1/4} \cdot |D_{1/a} b|^{-k} \left| \int \Delta^k [W^2(\|\zeta_a\|) V(\omega(\zeta_a)/\sqrt{a})] e^{-i(D_{1/a} b)'\zeta} d\zeta \right| \\ &\leq 2\pi \cdot a^{1/4} \cdot |D_{1/a} b|^{-k} \int \left| \Delta^k [W^2(\|\zeta_a\|) V(\omega(\zeta_a)/\sqrt{a})] \right| d\zeta. \end{aligned}$$

Hence

$$\begin{aligned} &(1 + |D_{1/a} b|^k) \cdot |\langle \gamma_{a,b,\theta}, \psi_{a_0,b_0} \rangle| \\ &\leq 2\pi \cdot a^{1/4} \int \left[ \left| W^2(\|\zeta_a\|) \right| \left| V(\omega(\zeta_a)/\sqrt{a}) \right| + \left| \Delta^k [W^2(\|\zeta_a\|) V(\omega(\zeta_a)/\sqrt{a})] \right| \right] d\zeta. \quad (9.2) \end{aligned}$$

Next we show that, for each $k$, there exists $c_k < \infty$ such that

$$\int \left[ \left| W^2(\|\zeta_a\|) \right| \left| V(\omega(\zeta_a)/\sqrt{a}) \right| + \left| \Delta^k [W^2(\|\zeta_a\|) V(\omega(\zeta_a)/\sqrt{a})] \right| \right] d\zeta \leq c_k, \quad \forall\, a > 0. \quad (9.3)$$

We have

$$\frac{\partial}{\partial \zeta_1} W^2(\|\zeta_a\|) = \frac{\partial}{\partial \zeta_1} W^2(\|(\,\cdot\,, \sqrt{a}\zeta_2)\|)(\zeta_1)$$

and

$$\frac{\partial}{\partial \zeta_2} W^2(\|\zeta_a\|) = \sqrt{a} \cdot \frac{\partial}{\partial \zeta_2} W^2(\|(\zeta_1, \sqrt{a}\,\cdot)\|)(\zeta_2).$$

Hence, by induction, the absolute values of the derivatives of $W^2(\|\zeta_a\|)$ are upper bounded independently of $a$. Also,

$$\frac{\partial}{\partial \zeta_1} V(\omega(\zeta_a)/\sqrt{a}) = \frac{\partial}{\partial \zeta_1} V(\omega((\,\cdot\,, \zeta_2)_a)/\sqrt{a})(\zeta_1) \cdot g_1(\zeta, a)$$

and

$$\frac{\partial}{\partial \zeta_2} V(\omega(\zeta_a)/\sqrt{a}) = \frac{\partial}{\partial \zeta_2} V(\omega((\zeta_1, \,\cdot\,)_a)/\sqrt{a})(\zeta_1) \cdot g_2(\zeta, a),$$

and tedious computations show that both $|g_1|, |g_2|$ possess an upper bound independently of $a$. Thus, by induction, the absolute values of the derivatives of $V(\omega(\zeta_a)/\sqrt{a})$ are upper bounded independently of $a$. These observations imply (9.3).

Further, for each $k = 1, 2, ..,$

$$\langle |D_{1/a} b| \rangle^k = (1 + |D_{1/a} b|^2)^{\frac{k}{2}} \leq \frac{k}{2}(1 + |D_{1/a} b|^k). \quad (9.4)$$

To finish, simply combine (9.2), (9.3), and (9.4), and recall that we chose coordinates so that $\theta = 0$. Translating back to the case of general $\theta$ gives the full conclusion. □

## 9.3 Proofs of Results from Section 4

### 9.3.1 Proof of Lemma 4.1

**Proof.** By Parseval,

$$\langle \psi_{a_{j'},b}, \mathcal{P}_j \rangle = 2\pi \int \hat{\psi}_{a_{j'},b}(\xi)\hat{\mathcal{P}}_j(\xi)d\xi = 2\pi \int a_{j'} W(a_{j'}r)e^{-ib'\xi} W(a_j r) r^{-1/2} d\xi,$$

where of course $r = |\xi|$. Now

$$W(a_{j'}r)W(a_j r) = 0 \qquad \forall r \geq 0, \qquad |j - j'| > 1,$$

hence we may as well assume that $j' = j$. Making the change of variables $\zeta = a_j\xi$, $d\zeta = a_j^2 d\xi$ and defining the annulus $\mathcal{A} = \{\zeta : 1/2 \leq |\zeta| \leq 2\}$,

$$\langle \psi_{a_j,b}, \mathcal{P}_j \rangle = 2\pi \cdot a_j^{-1/2} \cdot \int_{\mathcal{A}} W^2(|\zeta|)|\zeta|^{-1/2} e^{-i(b/a_j)'\zeta} d\zeta.$$

Applying integration by parts, for any $k = 0, 1, ...,$

$$\begin{aligned} |\langle \psi_{a_j,b}, \mathcal{P}_j \rangle| &= 2\pi \cdot a_j^{-1/2} \cdot |b/a_j|^{-k} \cdot \left| \int_{\mathcal{A}} \Delta^k [W^2(|\zeta|)|\zeta|^{-1/2}] e^{-i(b/a_j)'\zeta} d\zeta \right| \\ &\leq 2\pi \cdot a_j^{-1/2} \cdot |b/a_j|^{-k} \cdot \int_{\mathcal{A}} \left| \Delta^k [W^2(|\zeta|)|\zeta|^{-1/2}] \right| d\zeta. \end{aligned}$$

Hence

$$(1+|b/a_j|^k)\cdot|\langle \psi_{a_j,b}, \mathcal{P}_j \rangle| \leq 2\pi \cdot a_j^{-1/2} \cdot \int_{\mathcal{A}} \left[ \left| W^2(|\zeta|)|\zeta|^{-1/2} \right| + \left| \Delta^k [W^2(|\zeta|)|\zeta|^{-1/2}] \right| \right] d\zeta. \quad (9.5)$$

For $W$ suitably chosen,

$$\int_{\mathcal{A}} \left[ \left| W^2(|\zeta|)|\zeta|^{-1/2} \right| + \left| \Delta^k [W^2(|\zeta|)|\zeta|^{-1/2}] \right| \right] d\zeta < \infty.$$

Further, for each $k = 1, 2, ..,$

$$\langle |b/a_j| \rangle^k = (1 + |b/a_j|^2)^{\frac{k}{2}} \leq \frac{k}{2}(1 + |b/a_j|^k).$$

Infusing these two last observations into (9.5), for any $N = 1, 2, ...,$

$$|\langle \psi_{a_{j'},b}, \mathcal{P}_j \rangle| \leq c_N \cdot a_j^{-1/2} \cdot 1_{|j-j'|\leq 1} \cdot \langle |b/a_j| \rangle^{-N}.$$

□

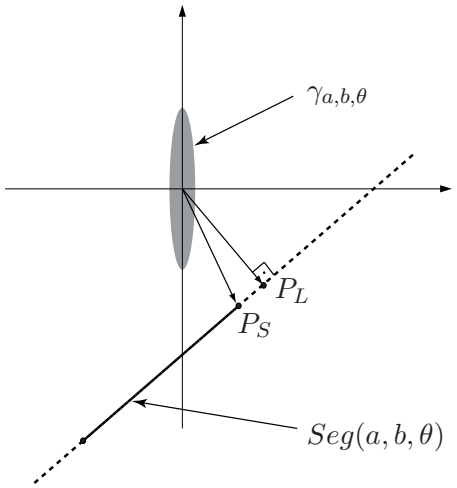


Figure 9: Relation between a curvelet $\gamma_{a,b,\theta}$, a line segment $Seg(a,b,\theta)$, its affine hull $Line(a,b,\theta)$, and the points $P_L$ and $P_S$.

## 9.4 Proofs of Results from Section 5

### 9.4.1 Proof of Lemma 5.1

We study the situation geometrically, and for each $(a,b,\theta)$ define the line segment

$$Seg(a,b,\theta) = \Big\{ D_{1/a} R_{-\theta} \begin{pmatrix} -b_1 \\ y - b_2 \end{pmatrix} : |y| \le \rho \Big\}.$$

Two special points associated to these line segments will play an essential role in our estimate; they are defined in

**Lemma 9.1** *Retain the definitions for $d_1, d_2, \sigma_1, \sigma_2$, and $\tau$ from the statement of Lemma 5.1. Then, for each $a, b, \theta$, the following conditions are fulfilled.*

(i) *Consider the line*

$$Line(a,b,\theta) = \Big\{ D_{1/a} R_{-\theta} \begin{pmatrix} -b_1 \\ y - b_2 \end{pmatrix} : y \in \mathbf{R} \Big\}.$$

*The closest point $P_L$ to the origin on $Line(a,b,\theta)$ satisfies*

$$\|P_L\|_2^2 = b_1^2(\sigma_2^2 - \sigma_1^{-2}\tau) = d_1^2.$$

(ii) *Let $P_S$ be the closest point on $Seg(a,b,\theta)$ to the origin. Then*

$$\|P_S - P_L\|_2^2 = d_2^2.$$

Figure 9 shows a general configuration featuring $P_L$ and $P_S$.

**Proof.** Set

$$L(y) := D_{1/a} R_{-\theta} \begin{pmatrix} -b_1 \\ y - b_2 \end{pmatrix}.$$

Then

$$\begin{aligned}\|L(y)\|_2^2 &= \|(a^{-1}(-b_1\cos\theta + (y-b_2)\sin\theta), a^{-1/2}(b_1\sin\theta + (y-b_2)\cos\theta)\|_2^2 \\ &= b_1^2\sigma_2^2 + (y-b_2)^2\sigma_1^2 + 2b_1(y-b_2)\tau. \end{aligned} \tag{9.6}$$

Since

$$\frac{d}{dy}\|L(y)\|_2^2 = 2(y-b_2)\sigma_1^2 + 2b_1\tau,$$

it follows by definition of $P_L$ that

$$P_L = L(b_2 - \sigma_1^{-2}b_1\tau). \tag{9.7}$$

Hence, by (9.6),

$$d_1^2 = \|P_L\|_2^2 = b_1^2\sigma_2^2 + (-\sigma_1^{-2}b_1\tau)^2\sigma_1^2 + 2b_1(-\sigma_1^{-2}b_1\tau)\tau = b_1^2(\sigma_2^2 - \sigma_1^{-2}\tau).$$

This proves (i).

To prove (ii), observe that, by (9.7), $P_L \in Seg(a,b,\theta)$ if and only if $b_2 - \sigma_1^{-2}b_1\tau \in [-\rho,\rho]$, which are the two different cases the definition of $d_2^2$ is separated into. Now, if $P_L \in Seg(a,b,\theta)$, then, obviously, $d_2^2 = 0$. Next assume that $P_L \notin Seg(a,b,\theta)$. Then

$$\begin{aligned} d_2^2 &= \min_{\pm} \|L(\pm\rho) - P_L\|_2^2 \\ &= \min_{\pm} (\pm\rho - b_2)^2\sigma_1^2 + \sigma_1^{-4}b_1^2\tau^2\sigma_1^2 + 2(\rho - b_2)\sigma_1^{-2}b_1\tau\sigma_1^2 \\ &= \min_{\pm} ((\pm\rho - b_2)\sigma_1 - \sigma_1^{-1}b_1\tau)^2. \end{aligned}$$

□

Now define the *ray integral*

$$R_N(x_0,y_0) = \int_{y_0}^{\infty} \langle|(x_0,t)|\rangle^{-N} dt,$$

i.e., we integrate along the vertical ray $\mathcal{R}(x_0,y_0)$ whose 'lowest' point is $(x_0,y_0)$. The geometry of the previous lemma allows to control the curvelet coefficient of a linear singularity by a ray integral, properly deployed. Farther below we will prove:

**Lemma 9.2** *Let*

$$\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle = \int_{-\rho}^{\rho} w(x_2/\rho)\gamma_{a,b,\theta}(0,x_2)dx$$

*where* $\|w\|_\infty \le 1$. *Then*

$$|\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| \le a^{-3/4}\cdot\sigma_1^{-1}\cdot R_N(d_1,\sigma_1 d_2).$$

The next lemma gives a bound on the ray integral which, combined with the last lemma, finishes the proof of Lemma 5.1.

**Lemma 9.3** *For* $y_0 \ge 0$,

$$R_N(x_0,y_0) \le \pi\cdot\langle|x_0|\rangle^{-1}\cdot\langle|(x_0,y_0)|\rangle^{2-N}. \tag{9.8}$$

**Proof of Lemma 9.3.** For $\beta \in (0,1)$,

$$\int_0^\infty |f(t)|dt \le (\sup_{t\in(0,\infty)} |f(t)|^\beta) \cdot \int_0^\infty |f(t)|^{1-\beta} dt.$$

Now setting $(1-\beta)N = 2$ and $f(t) = \langle|(x_0, y_0+t)|\rangle^{-N}$, we have

$$R_N(x_0,y_0) \le (\sup_{v\in\mathcal{R}(x_0,y_0)} \langle|v|\rangle^{2-N}) \cdot \int_0^\infty |\langle|(x_0,y_0+t)|\rangle^{-2} dt.$$

Since

$$\int_{-\infty}^\infty |\langle|(x_0,y)|\rangle^{-M} dy = \langle|x_0|\rangle^{-M} \cdot \int_{-\infty}^\infty \langle y/\langle x_0\rangle\rangle^{-M} dy = \langle|x_0|\rangle^{-M+1} \cdot \int_{-\infty}^\infty \langle t\rangle^{-M} dt,$$

setting $M=2$ and recalling $\pi = \int_{-\infty}^\infty (1+t^2)^{-1} dt$, it follows that

$$\int_0^\infty |\langle|(x_0,y_0+t)|\rangle^{-2} dt \le \pi \cdot \langle|x_0|\rangle^{-1}.$$

Meanwhile, since $y_0 \ge 0$,

$$\sup_{v\in\mathcal{R}(x_0,y_0)} \langle|v|\rangle^{2-N} = \langle|(x_0,y_0)|\rangle^{2-N}.$$

This proves (9.8). □

**Proof of Lemma 9.2.** By Lemma 3.2, Lemma 9.1, and using the fact that $\|w\|_\infty \le 1$,

$$\begin{aligned} |\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| &= \Big|\int_{-\rho}^{\rho} w_2(y/\rho)\gamma_{a,b,\theta}(0,y)dy\Big| \\ &\le \int_{-\rho}^{\rho} |\gamma_{a,b,\theta}(0,y)|dy \\ &\le \int_{Seg(a,b,\theta)} a^{-3/4}\langle|v|\rangle^{-N} dv \end{aligned}$$

where we used an affine transformation of variables to turn the anisotropic norm $|(0,y)|_{a,\theta}$ into the Euclidean norm $|v|$; the same transformation turns $\{0\}\times[-\rho,\rho]$ into $Seg(a,b,\theta)$. In the final expression, the integral is along a non-unit-speed curve traversing $Seg(a,b,\theta)$, at speed $\sigma_1$. Now let $Ray(a,b,\theta)$ denote the ray starting from $P_S$ and initially traversing $Seg(a,b,\theta)$. We continue with

$$\begin{aligned} c_N \cdot a^{-3/4} \cdot \int_{Seg(a,b,\theta)} \langle|v|\rangle^{-N} dv &\le c_N \cdot a^{-3/4} \cdot \int_{Ray(a,b,\theta)} \langle|v|\rangle^{-N} dv \\ &= c_N \cdot a^{-3/4} \cdot \sigma_1^{-1} \int_{\sigma_1\cdot Ray(a,b,\theta)} \langle|w|\rangle^{-N} dw \qquad (9.9) \\ &= c_N \cdot a^{-3/4} \cdot \sigma_1^{-1} \int_{\sigma_1 d_2}^\infty \langle|(d_1,t)|\rangle^{-N} dt \\ &= c_N \cdot a^{-3/4} \cdot \sigma_1^{-1} \cdot R_N(d_1, \sigma_1 d_2). \end{aligned}$$

In (9.9), the integral involves a unit-speed curve traversing $Ray(a,b,\theta)$, which explains the appearance of the speed factor $\sigma_1$. □

### 9.4.2 Proof of Lemma 5.2

By definition of the line singularity $w\mathcal{L}$ and by (5.1), we can rewrite $\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle$ in the following way:

$$2\pi\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle = \int (\hat{w} \star \hat{\gamma}_{a,b,\theta})(\xi_1, 0) d\xi_1. \tag{9.10}$$

Since

$$(\hat{w} \star \hat{\gamma}_{a,b,\theta})(\xi) = \int \rho \cdot \hat{w}_2(\rho(\xi_2 - \eta_2))\delta_0(\xi_1 - \eta_1)\hat{\gamma}_{a,b,\theta}(\eta) d\eta,$$

it follows that

$$(\hat{w} \star \hat{\gamma}_{a,b,\theta})(\xi_1, 0) \quad = \quad e^{ib_1\xi_1} \int \rho \cdot \hat{w}_2(-\rho\eta_2)\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2) e^{ib_2\eta_2} d\eta_2.$$

By (9.10), this implies

$$2\pi\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle = \int e^{ib_1\xi_1} \Big[ \int \rho \cdot \hat{w}_2(-\rho\eta_2)\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2) e^{ib_2\eta_2} d\eta_2 \Big] d\xi_1.$$

Repeatedly applying integration by parts, and incorporating analyst's brackets $\langle | \cdot | \rangle$ as in the proof of Lemma 4.1, we obtain

$$2\pi|\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| \leq \langle |b_1| \rangle^{-L} \cdot \langle |b_2| \rangle^{-M} \cdot \|h_{L,M}\|_{L^1(\mathbf{R})}, \tag{9.11}$$

where

$$h_{L,M}(\xi_1) = \rho \cdot \int D^{L,M}\Big( \hat{w}_2(-\rho\eta_2)\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2) e^{ib_2\eta_2} \Big) d\eta_2$$

and for some 'nice' $f \in L^2(\mathbf{R}^2)$,

$$D^{L,M} f(\eta_1, \eta_2) = \left( \frac{\partial}{\partial \eta_1} \right)^L \left( \frac{\partial}{\partial \eta_2} \right)^M f(\eta_1, \eta_2).$$

Next, we will estimate the term $|h_{L,M}(\xi_1)|$ from (9.11), and prove that

$$|h_{L,M}(\xi_1)| \leq c \cdot a^{3/4} \cdot e^{-\rho\frac{|\sin\theta|}{2a}} \cdot (a^{1/2}|\sin\theta| + a|\cos\theta|)^L \cdot (\rho + a^{1/2}|\cos\theta| + a|\sin\theta|)^M. \tag{9.12}$$

Let $\Xi_{a,\theta}(\xi_1)$ denote the support of the function $\xi_1 \mapsto D^{L,M}(\hat{w}(\rho\eta_2)\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2)e^{ib_2\eta_2})$. Then $h_{L,M}$ can be written as

$$h_{L,M}(\xi_1) = \rho \cdot \int_{\Xi_{a,\theta}(\xi_1)} D^{L,M}\Big( \hat{w}_2(-\rho\eta_2)\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2) e^{ib_2\eta_2} \Big) d\eta_2. \tag{9.13}$$

We next rewrite the integrand as

$$\begin{aligned} & D^{L,M}\Big( \hat{w}_2(-\rho\eta_2)\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2) e^{ib_2\eta_2} \Big) \\ & = \sum_{m=0}^{M} \binom{M}{m} \hat{w}_2^{(m)}(-\rho\eta_2)(-\rho)^m D^{L,M-m}(\hat{\gamma}_{a,0,\theta}(\xi_1, \eta_2) e^{ib_2\eta_2}). \end{aligned}$$

This allows us to estimate $|h_{L,M}(\xi_1)|$ using (9.13) and (6.2) by

$$\begin{aligned}
|h_{L,M}(\xi_1)| &\leq \sum_{m=0}^{M} \binom{M}{m} \Big| \int_{\Xi_{a,\theta}(\xi_1)} \hat{w}_2^{(m)}(-\rho\eta_2)(-\rho)^m D^{L,M-m}(\hat{\gamma}_{a,0,\theta}(\xi_1,\eta_2)e^{ib_2\eta_2})d\eta_2 \Big| \\
&\leq \sum_{m=0}^{M} \binom{M}{m} \rho^{m+1} \cdot \|\hat{w}_2^{(m)}(\rho\,\cdot)\|_{L^1[|\sin\theta|/(2a),\infty)} N^{L,M-m}(a,\theta) \\
&\leq c \cdot \sum_{m=0}^{M} \binom{M}{m} \rho^{m+1} \cdot e^{-\rho\frac{|\sin\theta|}{2a}} \cdot N^{L,M-m}(a,\theta), \qquad (9.14)
\end{aligned}$$

where

$$N^{L,M-m}(a,\theta) = \|D^{L,M-m}\hat{\gamma}_{a,0,\theta}(\xi_1,\eta_2)\|_{L^\infty(\Xi_{a,\theta}(\xi_1))}.$$

Since, by simple decay estimates,

$$|D_1^L\hat{\gamma}_{a,0,\theta}(\eta_1,\eta_2)| \leq C_L \cdot a^{3/4} \cdot (a|\cos\theta| + a^{1/2}|\sin\theta|)^L$$

and

$$|D_2^M\hat{\gamma}_{a,0,\theta}(\eta_1,\eta_2)| \leq C_M \cdot a^{3/4} \cdot (a^{1/2}|\cos\theta| + a|\sin\theta|)^M,$$

the term $N^{L,M-m}(a,\theta)$ can be estimated by

$$N^{L,M-m}(a,\theta) \leq c_{L,M} \cdot a^{3/4} \cdot (a|\cos\theta| + a^{1/2}|\sin\theta|)^L (a^{1/2}|\cos\theta| + a|\sin\theta|)^M.$$

Combining this finding with (9.14) proves (9.12).

Thus, in particular, by the support of the function $h_{L,M}$, the $L^1$-norm of this function can be estimated as

$$\begin{aligned}
&\|h_{L,M}\|_{L^1(\mathbf{R})} \\
&\leq c \cdot a^{-1} \cdot |\cos\theta| \cdot a^{3/4} \cdot e^{-\rho\frac{|\sin\theta|}{2a}} \cdot (a^{1/2}|\sin\theta| + a|\cos\theta|)^L \cdot (\rho + a^{1/2}|\cos\theta| + a|\sin\theta|)^M.
\end{aligned}$$

Combining this estimate with (9.11) yields

$$\begin{aligned}
|\langle w\mathcal{L}, \gamma_{a,b,\theta}\rangle| &\leq c_{M,L} \cdot a^{-1/4} \cdot |\cos\theta| \cdot e^{-\rho\frac{|\sin\theta|}{2a}} \cdot \langle|b_1|\rangle^{-L} \cdot (a^{1/2}|\sin\theta| + a|\cos\theta|)^L \\
&\quad \cdot \langle|b_2|\rangle^{-M} \cdot (\rho + a^{1/2}|\cos\theta| + a|\sin\theta|)^M,
\end{aligned}$$

as claimed. □

## 9.5 Proofs of Results from Section 7

### 9.5.1 Proof of Lemma 7.8

**Proof.** Below, various constants will appear, which for simplicity shall all be denoted by $c$. Also we write $b_{\tilde{k}}$, $b_{k'}$, etc. in place of the full notation $b_{\tilde{j},\tilde{k},\tilde{\ell}} = R_{\theta_{j,\ell}} D_{2^{-j}} k$, etc. Throughout the proof we associate $\tilde{\eta}$ with the triple $(\tilde{j},\tilde{k},\tilde{\ell})$ and similarly for $\eta'$ and $(j',k',\ell')$.

Suppose first that $j' \geq \tilde{j}$. Then

$$\begin{aligned}
&\omega(\eta', \tilde{\eta}) \\
&\quad = \quad 2^{|\tilde{j}-j'|}\left(1 + \min\{2^{\tilde{j}}, 2^{j'}\}\left[|\theta_{\tilde{j},\tilde{\ell}} - \theta_{j',\ell'}|^2 + |b_{\tilde{k}} - b_{k'}|^2 + |\langle e_{\eta'}, b_{\tilde{k}} - b_{k'}\rangle|\right]\right) \\
&\quad \leq \quad c \cdot \left(2^{j'-\tilde{j}} + |\tilde{\ell} - 2^{(\tilde{j}-j')/2}\ell'|^2 + 2^{\tilde{j}} \cdot |b_{\tilde{k}} - b_{k'}|^2 + 2^{\tilde{j}}|\langle e_{\eta'}, b_{\tilde{k}} - b_{k'}\rangle|\right).
\end{aligned}$$

Similarly, if $j' < \tilde{j}$:

$$\omega(\eta', \tilde{\eta}) \quad \leq \quad c \cdot \left(2^{\tilde{j}-j'} + |2^{(j'-\tilde{j})/2}\tilde{\ell} - \ell'|^2 + 2^{j'} \cdot |b_{\tilde{k}} - b_{k'}|^2 + 2^{j'}|\langle e_{\eta'}, b_{\tilde{k}} - b_{k'}\rangle|\right). \quad (9.15)$$

Using the model (7.6), we now seek to prove that, for sufficiently large $j$,

$$\#\{\eta' : |\tilde{M}_j^i(\eta', \eta)| \geq c_N \cdot n_j^{-N}\} = \#\{\eta' : \omega(\eta', (\tilde{j}, \tilde{k}, \tilde{\ell})) \leq n_j\} \geq n_j. \quad (9.16)$$

We first study the case $j' < \tilde{j}$. Let $\tilde{\omega}(\eta', \tilde{\eta})$ denote the RHS of (9.15). Note that, if we can prove (9.16) with $\tilde{\omega}(\eta', \tilde{\eta})$ in place of $\omega(\eta', \tilde{\eta})$, this immediately implies the original claim of (9.16). We repeatedly use the (trivial)

**Lemma 9.4** *For each fixed $x \in \mathbf{R}$, the set of $k \in \mathbf{Z}$ satisfying $|k - x| \leq R$ has cardinality $\geq R - 1$.*

Define $C_{j,j',\tilde{j}} = n_j/c - 2^{j'-\tilde{j}}$, where $c$ is the constant in (9.15). The condition $\tilde{\omega} \leq n_j$ is equivalent to

$$\|2^{(j'-\tilde{j})/2}\tilde{\ell} - \ell'|^2 + 2^{j'} \cdot |b_{\tilde{k}} - b_{k'}|^2 + 2^{j'}|\langle e_{\eta'}, b_{\tilde{k}} - b_{k'}\rangle| \quad \leq \quad C_{j,j',\tilde{j}}. \quad (9.17)$$

We next derive conditions making each of the three terms smaller than $C_{j,j',\tilde{j}}/3$, thus implying (9.17). Fix $j'$. By Lemma 9.4, there are at least $C_{j,j',\tilde{j}}^{1/2}/\sqrt{3} - 1$ integer values $\ell' \in \mathbf{Z}$ obeying

$$|2^{(j'-\tilde{j})/2}\tilde{\ell} - \ell'|^2 \leq \frac{1}{3}C_{j,j',\tilde{j}}. \quad (9.18)$$

Now, define $\tilde{x} = \tilde{x}(\tilde{\eta}, \eta') = R_{\theta_{j',\ell'}}^{-1} b_{\tilde{k}}$. Since $R_{\theta_{j',\ell'}}$ is an isometry,

$$|b_{\tilde{k}} - b_k|^2 = |\tilde{x} - D_{2^{-j'}}k'|^2.$$

At the same time,

$$|\langle e_{\eta'}, b_{\tilde{k}} - b_{k'}\rangle| = |\tilde{x}_1 - 2^{-j'}k_1'|.$$

Now there are at least $C_{j,j',\tilde{j}}^{1/2}/\sqrt{6} - 1$ integers $k_2' \in \mathbf{Z}$ obeying

$$|2^{j'/2}\tilde{x}_2 - k_2'| \leq C_{j,j',\tilde{j}}^{1/2}/\sqrt{6}. \quad (9.19)$$

For large $j'$, there are at least $C_{j,j',\tilde{j}}/3 - 1$ integers $k_1' \in \mathbf{Z}$ obeying *both*

$$|2^{j'}\tilde{x}_1 - k_1'| \leq 2^{j'/2} \cdot C_{j,j',\tilde{j}}^{1/2}/\sqrt{6}$$

*and*

$$|2^{j'}\tilde{x}_1 - k'_1| \le C_{j,j',\tilde{j}}/3. \tag{9.20}$$

Every pair $k' = (k'_1, k'_2) \in \mathbf{Z}^2$ satisfying the conditions (9.19)-(9.20) simultaneously, satisfies

$$|b_{\tilde{k}} - b_k|^2 \le C_{j,j',\tilde{j}}/3 \quad \text{and} \quad |\langle e_{\eta'}, b_{\tilde{k}} - b_{k'}\rangle| \le C_{j,j',\tilde{j}}/3. \tag{9.21}$$

Combining the above displays, we have at least $(C_{j,j',\tilde{j}}/3 - 1) \cdot (C^{1/2}_{j,j',\tilde{j}}/\sqrt{6} - 1)$ points $k' = (k'_1, k'_2) \in \mathbf{Z}^2$ satisfying (9.21). Every pair $(k', \ell')$ satisfying (9.21) and (9.18) satisfies (9.17). So focusing just on $j' = j$, we obtain a large number of pairs $(k', \ell')$ satisfying (9.17): $\ge c \cdot C^2_{j,j',\tilde{j}} \sim c \cdot n_j^2 \gg n_j$ such pairs. In particular, for all large $j$, inequality $\omega(\tilde{\eta}, \eta') \le n_j$ is satisfied by at least $n_j$ triples $\eta' = (j', k', \ell')$.

For the case $j' < \tilde{j}$, using (9.15), we can similarly prove that $\omega \le n_j$ is satisfied by at least $n_j$ triples $(j', k', \ell')$.

Finally, we observe that $\omega(\tilde{\eta}, \eta') \le n_j$ can only hold if

$$|j' - \tilde{j}| \le c \cdot \log n_j, \quad |2^{(j'-\tilde{j})/2}\tilde{\ell} - \ell'| \le c \cdot n_j \tag{9.22}$$

and

$$2^{j'} \cdot \max\{|b_{\tilde{k}} - b_{k'}|^2, |b_{\tilde{k}} - b_{k'}|\} \le c \cdot n_j. \tag{9.23}$$

Concluding, each $\eta'$ contained in the sets in (9.16) must satisfy both (9.22) and (9.23). □

### 9.5.2 Proof of Lemma 7.9

For $\eta \in \tilde{\mathcal{S}}_j$, we have $\theta = 0$ and $|b_2| < n_j/2^j$ while $|b_1| < 2\rho/2^{j/2}$. WLOG suppose that, for the patch $i$ in question, we have that the $\eta' \in S^i_j$ of interest has $b' = R_{\theta'} D_{2^{-j'}} k'$ for some $\theta$ obeying $|\theta| \le cn_j/2^{j/2}$. For such a pair $(\eta, \eta')$, note that

$$k = D_{2^j} R_0^{-1} b, \qquad k' = D_{2^{j'}} R_{\theta'}^{-1} b';$$

it is also convenient to define $x' = D_{2^{j-j'}} k'$. Then

$$k_2 - x'_2 = 2^{j/2}\left((1 - \cos(\theta'))b'_2 + (b_2 - b'_2) - \sin(\theta')b'_1\right).$$

Now $|b'_1| < C$ for $\eta' \in \text{FWD}(\eta)$, and, from Lemma 7.8 we infer

$$|1 - \cos(\theta')| < cn_j^2/2^j, \qquad |b'_2 - b_2| < cn_j/2^j, \qquad |\sin(\theta')| < cn_j/2^{j/2}.$$

Combining these,

$$|k_2 - x'_2| < 2cn_j^2,$$

so $|x'_2| > |k_2| - 2cn_j^2$, and

$$|k'_2| = 2^{(j'-j)/2}|x'_2| \ge 2^{-|j-j'|/2} \cdot \left(|k_2| - 2cn_j^2\right).$$

Now $2^{|j-j'|} < cn_j$ for $\eta' \in \text{FWD}(\eta)$, so $|k(\eta)| \le |k_2(\eta)| + |k_1(\eta)| \le |k_2(\eta)| + cn_j$, for $\eta \in \tilde{\mathcal{S}}_j$. So

$$|k(\eta')| \ge |k'_2| \ge 2^{-|j-j'|/2} \cdot \left(|k_2| - 2cn_j^2\right), \qquad j > j_0.$$

□

### 9.5.3 Proof of Lemma 7.10

Let $Q_\ell$ denote the square of sidelength $2^\ell + 1$ centered at the origin. Within each annulus $A_\ell = Q_\ell - Q_{\ell-1}$ there are, for $\ell \geq 2$, fewer than $2^{2\ell}$ points in the annulus and each one has $\ell^2$ norm at least $2^{\ell-2}$. Partition the sum $\sum_{\mathbf{Z}^2} = \sum_{Q_1+A_2+A_3+\ldots}$. Starting at $\ell = 2$ each annulus contributes at most $2^{2\ell} \cdot \langle (a2^{\ell-2} - b)_+ \rangle^{-N}$ to the sum, and the inner square $Q_1$ contributes at most 9. Then

$$\begin{aligned} \sum_{k \in \mathbf{Z}^2} \langle (a|k| - b)_+ \rangle^{-N} &\leq 9 + \sum_{m=2}^{\infty} 2^{2\ell} \cdot \langle (a2^{\ell-2} - b)_+ \rangle^{-N} \\ &= 9 + 16 \cdot \left( \sum_{m=0}^{\infty} 2^{2m} \cdot \langle (a2^m - b)_+ \rangle^{-N} \right). \end{aligned}$$

Let $m_0$ satisfy $2b \geq a2^{m_0} \geq b$. Now

$$T_1 := \sum_{m=0}^{m_0} 2^{2m} \langle |a2^m - b|_+ \rangle^{-N} \leq \sum_{m=0}^{m_0} 2^{2m} \leq 2^{2m_0+1}.$$

Now $2^{2m_0} \leq (2b/a)^2$, so $2^{m_0} \leq 2b/a$. So $T_1 \leq 2 \cdot (2b/a)^2$.

On the other hand, $a2^m - b \geq 0$ for $m \geq m_0$, while $a2^{m-m_0} \leq a2^m - b$. Then

$$\begin{aligned} T_2 := \sum_{m > m_0} 2^{2m} \langle |a2^{m-m_0}| \rangle^{-N} &\leq \sum_{m=m_0+1}^{\infty} 2^{2m} (a2^{m-m_0})^{-N} \\ &\leq 2^{2(m_0+1)} (a2^{-m_0})^N \sum_{h=0}^{\infty} 2^{2h} 2^{-hN} \\ &= 2^{m_0(2-N)} \cdot a^N \cdot 2 \cdot (1 - 2^{-(N-2)})^{-1} \\ &\leq (b/a)^{2-N} \cdot a^N \cdot 4 \\ &\leq 4 \cdot (2b/a)^2 \cdot b^{-N} \end{aligned}$$

Hence

$$T_1 + T_2 \leq (b/a)^2 \cdot \left(8 + 16 b^{-N}\right).$$

Combining these displays gives the lemma, with explicit constants. □

# References


[1] J. Bobin, J.-L. Starck, M.J. Fadili, Y. Moudden, and D.L. Donoho, *Morphological Component Analysis: An Adaptive Thresholding Strategy*, IEEE Trans. Image Proc. **16(11)** (2007), 2675–2681.

[2] L. Borup, R. Gribonval, and M. Nielsen, *Beyond Coherence : Recovering Structured Time-Frequency Representations*, Appl. Comput. Harmon. Anal. **24(1)** (2008), 120–128.

[3] A.M. Bruckstein, D.L. Donoho, and M. Elad, *From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images,* SIAM Review **51(1)** (2009), 34–81.

[4] F. Bunea, A. Tsybakov and M. Wegkamp, Aggregation for Gaussian Regression, Annals of Statistics **35(4)** (2007), 1674–1697.

[5] E. J. Candès and L. Demanet, *The curvelet representation of wave propagators is optimally sparse,* Comm. Pure Appl. Math. **58(11)** (2005), 1472–1528.

[6] E. J. Candès and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with $C2$ singularities*, Comm. Pure Appl. Math. **56(2)** (2004), 219–266.

[7] E. J. Candès and D. L. Donoho, *Continuous curvelet transform: I. Resolution of the wavefront set*, Appl. Comput. Harmon. Anal. **19(2)** (2005), 162–197.

[8] E. J. Candès and D. L. Donoho, *Continuous curvelet transform: II. Discretization of frames*, Appl. Comput. Harmon. Anal. **19(2)** (2005), 198–222.

[9] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math. **59(8)** (2006), 1207–1223.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit,* SIAM Rev. **43** (2001), 129–159.

[11] O. Christensen, *An introduction to frames and Riesz bases*, Birkhäuser, Boston, 2003.

[12] R. R. Coifman and M. V. Wickerhauser, *Wavelets and adapted waveform analysis. A toolkit for signal processing and numerical analysis*, Different perspectives on wavelets (San Antonio, TX, 1993), 119–153, Proc. Sympos. Appl. Math., **47**, Amer. Math. Soc., Providence, RI, 1993.

[13] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52(4)** (2006), 1289–1306.

[14] D. L. Donoho, *For most large underdetermined systems of linear equations the minimal $l_1$-norm solution is also the sparsest solution*, Comm. Pure Appl. Math. **59(6)** (2006), 797–829.

[15] D. L. Donoho, *For most large underdetermined systems of equations, the minimal $l_1$-norm near-solution approximates the sparsest near-solution*, Comm. Pure Appl. Math. **59(7)** (2006), 907–934.

[16] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via $l^1$ minimization*, Proc. Natl. Acad. Sci. USA **100(5)** (2003), 2197–2202.

[17] D. L. Donoho, M. Elad, and V. N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory **52(1)** (2006), 6–18.

[18] D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition,* IEEE Trans. Inform. Theory **47(7)** (2001), 2845–2862.

[19] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, *Maximum entropy and the nearly black object*, J. Roy. Statist. Soc. Ser. B **54(1)** (1992), 41–81.

[20] D. L. Donoho and G. Kutyniok, *Geometric Separation by Single Pass Alternating Thresholding*, manuscript (2010).

[21] D. L. Donoho and G. Kutyniok, *Sparsity Equivalence of Anisotropic Decompositions*, manuscript (2010).

[22] D. L. Donoho and B. F. Logan, *Signal recovery and the large sieve,* SIAM J. Appl. Math. **52(2)** (1992), 577–591.

[23] D. L. Donoho and P. B. Stark, *Uncertainty principles and signal recovery,* SIAM J. Appl. Math. **49(3)** (1989), 906–931.

[24] M. Elad and A. M. Bruckstein, *A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases,* IEEE Trans. Inform. Theory **48(9)** (2002), 2558–2567.

[25] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Appl. Comput. Harmon. Anal. **19(3)** (2005), 340–358.

[26] R. Gribonval and E. Bacry, *Harmonic decomposition of audio signals with matching pursuit,* IEEE Trans. Signal Proc. **51(1)** (2003), 101–111.

[27] R. Gribonval and M. Nielsen, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory **49(12)** (2003), 3320–3325.

[28] K. Guo, G. Kutyniok, and D. Labate, *Sparse Multidimensional Representations using Anisotropic Dilation und Shear Operators,* in Wavelets und Splines (Athens, GA, 2005), G. Chen und M. J. Lai, eds., Nashboro Press, Nashville, TN (2006), 189–201.

[29] L. Hörmander, *The analysis of linear partial differential operators. I. Distribution theory and Fourier analysis,* Springer-Verlag, Berlin, 2003.

[30] M. Kowalski and B. Torrésani, *Sparsity and Persistence: mixed norms provide simple signal models with dependent coefficients*, Signal, Image and Video Processing, to appear.

[31] G. Kutyniok and D. Labate, *Resolution of the Wavefront Set using Continuous Shearlets,* Trans. Amer. Math. Soc. **361(5)** (2009), 2719–2754.

[32] G. Kutyniok and T. Sauer, *Adaptive Directional Subdivision Schemes and Shearlet Multiresolution Analysis*, SIAM J. Math. Anal. **41(4)** (2009), 1436–1471.

[33] G. Kutyniok, M. Shahram, and D. L. Donoho, *Development of a Digital Shearlet Transform Based on Pseudo-Polar FFT*, in Wavelets XIII (San Diego, CA, 2009), D. Van De Ville, V. K. Goyal und M. Papadakis, eds., 74460B-1 - 74460B-13, SPIE Proc. **7446**, SPIE, Bellingham, WA, 2009.

[34] S. G. Mallat and Z. Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Proc. **41(12)** (1993), 3397–3415.

[35] F. G. Meyer, A. Averbuch, and R. R. Coifman, *Multi-layered Image Representation: Application to Image Compression*, IEEE Trans. on Image Processing **11(9)** (2002), 1072–1080.

[36] Y. Meyer (2001) *Oscillating patterns in image processing and nonlinear evolution equations.* Providence: Amer. Math. Soc.

[37] L. Vese and S. Osher (2003). *Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing.* Journal of Scientific Computing **19**, 553–572.

[38] H. F. Smith, *A Hardy space for Fourier integral operators*, J. Geom. Anal. **8**, 629–653.

[39] J.-L. Starck, E. Candès, and D. L. Donoho, *Astronomical Image Representation by the Curvelet Tansform*, Astronomy and Astrophysics **398** (2003), 785–800.

[40] J.-L. Starck, M. Elad, and D. L. Donoho, *Redundant Multiscale Transforms and their Application for Morphological Component Analysis*, Journal of Advances in Imaging and Electron Physics **132** (2004), 287–348.

[41] J.-L. Starck, M. Elad, and D. L. Donoho, *Image decomposition via the combination of sparse representations and a variational approach,* IEEE Trans. Image Proc. **14(10)** (2005), 1570–1582.

[42] J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D.L. Donoho, *Morphological Component Analysis*, Wavelets XI (San Diego, CA, 2005), SPIE Proc. 5914, SPIE, Bellingham, WA, 2005.

[43] J.-L. Starck, M. Nguyen, and F. Murtagh, *Wavelets and curvelets for image deconvolution: A combined approach,* Signal Process. **83** (2003), 2279–2283.

[44] G. Teschke, *Multi-frame representations in linear inverse problems with mixed multi-constraints*, Appl. Comput. Harmon. Anal. **22(1)** (2007), 43–60.

[45] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory **50(10)** (2004), 2231–2242.

[46] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, Journal of The Royal Statistical Society Series B **68(1)** (2006).

[47] M. Zibulevsky and B. Pearlmutter, *Blind source separation by sparse decomposition in a signal dictionary,* Neur. Comput. **13** (2001), 863–882.